

**Preprint No. M 10/10**

**Zur Numerik nichtlinearer  
dynamischer Systeme**

**Teil 1: Gleichgewichtslösungen**

Vogt, Werner

Oktober 2010

**Impressum:**

Hrsg.: Leiter des Instituts für Mathematik  
Weimarer Straße 25  
98693 Ilmenau  
Tel.: +49 3677 69-3621  
Fax: +49 3677 69-3270  
<http://www.tu-ilmenau.de/math/>

**Zusammenfassung** Der Beitrag stellt die wesentlichen Algorithmen vor, die der Berechnung und Analyse von Gleichgewichtslösungen nichtlinearer dynamischer Systeme – beschrieben durch gewöhnliche Differenzialgleichungen – dienen. Schwerpunkt sind dabei parameterabhängige Systeme, für die praktikable Fortsetzungsmethoden eingeführt werden. Schließlich werden auftretende Bifurkationen klassifiziert und Verfahren zu deren Detektierung vorgestellt. Zahlreiche Beispiele und Aufgaben dienen dabei der besseren Veranschaulichung abstrakter Sachverhalte.

**MSC 2010:** 37M20, 65P30, 65P40, 65H10

**Keywords:** Computational methods for bifurcation problems, nonlinear stabilities, systems of equations

## 1 Einleitung

Dynamische Systeme beschreiben als mathematische Objekte die zeitliche Entwicklung existierender Systeme in Natur-, Ingenieurs- und anderen Wissenschaften. Zeitkontinuierliche Systeme werden dabei häufig durch gewöhnliche Differenzialgleichungen (DGL) dargestellt. Wegen der Nichtlinearität der mathematischen Modelle ist allerdings in den meisten praktischen Anwendungen eine geschlossene Lösung der DGL nicht möglich und leistungsfähige numerische Verfahren müssen eingesetzt werden. Deshalb sind Naturwissenschaftler und Ingenieure auf effiziente Algorithmen und Computer bei der Erforschung der Modelle ihres Fachgebietes angewiesen. Für umfangreiche Computersimulationen realer Prozesse werden schnelle und zuverlässige numerische Verfahren benötigt, deren grundlegendes Verständnis und Einordnung in zunehmendem Maße Bestandteil der mathematischen Ausbildung werden sollte.

Wir wollen uns mit numerischen Verfahren für folgende Problemklasse befassen: Gegeben ist ein parameterabhängiges  $C^r$ -glattes Vektorfeld  $f \in C^r(\mathbb{R}^{n+1})$  mit  $r \geq 1$ , mit dem das autonome System

$$\dot{x} = f(x, \lambda), \quad f : D \times \mathbb{R} \rightarrow \mathbb{R}^n, \quad D \subset \mathbb{R}^n, \text{ offen}, \quad \dot{x} \equiv \frac{dx}{dt} \quad (1)$$

definiert wird. Dabei ist  $\lambda \in \Lambda$  ein reeller Parameter aus einem vorgegebenen endlichen Intervall  $\Lambda = [a, b]$ . Realistische dynamische Systeme besitzen meist eine Vielzahl von Parametern, die wir der Einfachheit wegen bis auf einen einzigen Parameter  $\lambda$ , den sogenannten *Systemparameter* (auch Kontrollparameter genannt), konstant halten wollen.

### Beispiel 1

1. W. F. Langford (1984) betrachtet ein dynamisches System der Fluidodynamik, das nach mehreren Problemtransformationen auf das DGL-System

$$\begin{aligned}
\dot{x}_1 &= (x_3 - 0.7) \cdot x_1 - \omega \cdot x_2 \\
\dot{x}_2 &= \omega x_1 + (x_3 - 0.7) \cdot x_2 \\
\dot{x}_3 &= 0.6 + x_3 - x_3^3/3 - (x_1^2 + x_2^2)(1 + \rho \cdot x_3) + \varepsilon \cdot x_3 \cdot x_1^3
\end{aligned} \tag{2}$$

mit den positiven reellen Parametern  $\omega, \rho$  und  $\varepsilon$  führt. Mit festen Werten  $\omega = 3.5$ ,  $\varepsilon = 0.03$  und dem Systemparameter  $\lambda = \rho$  ergibt sich die Darstellung (1).

2. Der Meteorologe E. N. Lorenz untersuchte bereits in den 1960er Jahren ein vereinfachtes Modell von Grundgleichungen (vgl. Argyris et al. (1995))

$$\begin{aligned}
\dot{x}_1 &= x_2 - x_1 \\
\dot{x}_2 &= \lambda x_1 - x_2 - x_1 x_3 \\
\dot{x}_3 &= x_1 x_2 - x_3
\end{aligned} \tag{3}$$

mit dem positiven Kontrollparameter  $\lambda$ . Wie (2) zeigt auch dieses System trotz scheinbar einfacher Struktur ein überaus reichhaltiges Lösungsverhalten. ■

Nachfolgend wollen wir *Gleichgewichtslösungen* parameterabhängiger Systeme (1) berechnen, die zugleich Fixpunkte des Flusses darstellen. Die dabei gewonnenen numerischen Techniken werden sich auch für weitere Lösungstypen, wie periodische und quasi-periodische Lösungen, als grundlegend erweisen.

## 2 Berechnung von Gleichgewichtslösungen

Punkte  $x^* \in D$  des Phasenraumes, in denen keine Bewegung stattfindet, bezeichnet man als Gleichgewichtspunkte oder Fixpunkte der DGL. Ihre numerische Bestimmung beruht auf folgender

### Definition 2 (Gleichgewichtspunkt, Attraktor, Repellor)

- i. Ein Punkt  $x^* \in D$  heißt *Gleichgewichtspunkt*, *Fixpunkt* oder *kritischer Punkt* der DGL (1) zu festem Parameterwert  $\lambda^*$ , wenn  $f(x^*, \lambda^*) = 0$  ist.
- ii. Der Gleichgewichtspunkt  $x^*$  ist ein (positiver) *Attraktor*, falls eine offene Umgebung  $\mathcal{U}(x^*) \subset D$  von  $x^*$  existiert, so dass für jede Lösung  $x(t)$  von (1) mit  $x(0) \in \mathcal{U}(x^*)$  stets  $\lim_{t \rightarrow \infty} x(t) = x^*$  folgt. Gilt diese Eigenschaft für  $t \rightarrow -\infty$ , so ist  $x^*$  ein *Repellor* (auch negativer Attraktor genannt).

Ist  $x^*$  ein Gleichgewichtspunkt zu festem Parameterwert  $\lambda^*$  und  $\varphi_t : D \rightarrow \mathbb{R}^n$  die Abbildung des Flusses der DGL (1), so ist offenbar  $\varphi_t(x^*) = x^*$  für alle  $t \in \mathbb{R}$ . Damit ist

$x^*$  auch ein *Fixpunkt des Flusses*, und die zugehörige konstante Lösung  $x(t) = x^*$  wird als *Gleichgewichtslösung* bzw. *Ruhelage* des dynamischen Systems bezeichnet.

Aus dem Existenz- und Eindeutigkeitssatz für die DGL (1) ergibt sich, dass ein Gleichgewichtspunkt mit  $x(0) \neq x^*$  niemals in endlicher Zeit erreicht werden kann. Deshalb wird man zu seiner Bestimmung eine Integration der DGL unbedingt vermeiden, insbesondere wenn  $x^*$  kein Attraktor bzw. Repellor ist. Wir gehen vielmehr direkt von Definition 2 aus und halten vorerst den Kontrollparameter  $\lambda^* \in \Lambda = [a, b]$  fest. Damit sind Gleichgewichtspunkte nun als Lösungen eines parameterlosen Gleichungssystems

$$f(x) = 0, \quad f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n \quad (4)$$

gesucht. Der Einfachheit halber werde vorausgesetzt, dass  $f$  auf  $D$  mindestens zweimal stetig differenzierbar ist.

## 2.1 Newton-Verfahren

Der bekannteste Lösungsansatz besteht darin, die Funktion  $f$  an einer zuvor ermittelten Näherungslösung  $x_k$  zu linearisieren. Entwickeln wir dazu  $f(x)$  an der Stelle  $x_k$  in die Taylor-Reihe

$$f(x) = f(x_k) + f'(x_k)(x - x_k) + R(x, x_k),$$

so liefert die rechte Seite bei Vernachlässigung des Restgliedes  $R(x, x_k)$  die *Linearisierung der Funktion*  $f$  am Näherungspunkt  $x_k$ . Anstelle der Gleichung  $f(x) = 0$  lösen wir nun die linearisierte Gleichung mit der Lösung  $x = x_{k+1}$

$$f(x_k) + f'(x_k)(x_{k+1} - x_k) = 0,$$

und erhalten nach Umstellung das *Newton-Verfahren*<sup>1</sup> (engl. *Newton-Raphson method*)

$$x_{k+1} = x_k - [F(x_k)]^{-1} f(x_k), \quad k = 0, 1, 2, 3, \dots \quad (5)$$

mit der Jacobi-Matrix

$$F(x) = f'(x) = \left( \frac{\partial f_i}{\partial x_j} \right). \quad (6)$$

Das Verfahren nimmt eine iterative Linearisierung der gegebenen Funktion an den berechneten Näherungswerten  $x_k$  vor. Dieser einfache Zugang, im eindimensionalen Falle auch als *Tangenten-Näherungsverfahren* bekannt, bildet zugleich eines der grundlegendsten und leistungsfähigsten Näherungsverfahren für differenzierbare Abbildungen.

---

<sup>1</sup>Sir Isaac Newton (1643–1727), engl. Physiker, Mathematiker und Astronom; Entdeckung der Gravitation; Newtonsche Axiome; Grundlagen der Differenzial- und Integralrechnung.

Bevor wir seine Konverenzeigenschaften näher betrachten, wollen wir eine effiziente algorithmische Darstellung angeben. Die Implementation des Newton-Verfahrens erfordert die Berechnung von  $f(x_k)$  und der Jacobi-Matrix  $F(x_k)$  sowie die Bestimmung der Näherung  $x_{k+1}$ . Nehmen wir an, die Jacobi-Matrix sei voll besetzt, und direkte Verfahren werden zur Lösung der linearen Gleichungssysteme eingesetzt. In jedem Schritt des Newton-Verfahrens ist dann die Invertierung der Jacobi-Matrix  $F(x_k)$  erforderlich. Der beträchtliche arithmetische Aufwand lässt sich jedoch auf  $\frac{1}{3}$  reduzieren, wenn man stattdessen in jedem Schritt ein lineares Gleichungssystem löst. Das erreichen wir, wenn wir die Verfahrensgleichung (5) nach der Newton-Korrektur  $d_k = x_{k+1} - x_k$  umstellen und mit der Jacobi-Matrix  $F(x_k)$  multiplizieren. So entsteht eine *praktikablere Form des Newton-Verfahrens*:

Löse für  $k = 0, 1, 2, \dots$  das lineare Gleichungssystem

$$F(x_k) d_k = -f(x_k), \quad x_{k+1} = x_k + d_k. \quad (7)$$

Der *arithmetische Aufwand* pro Newton-Schritt besteht nun in der Lösung eines linearen Gleichungssystems mit  $n$  Unbekannten, wofür die LU-Zerlegung oder QR-Zerlegung eingesetzt werden kann (vgl. Hoffmann et al. (2005)). Dafür sind  $\mathcal{O}(n^3)$  Gleitpunktoperationen (*floating point operations, flops*) erforderlich. Allerdings sind die  $n$  Funktionswerte

### ALGORITHMUS 3 (Newton-Verfahren)

Function  $[x] = \text{newton}(f, F, x_0, \text{tolabs}, \text{tolrel})$

1. Berechne  $x = x_0$ ,  $\text{tol} = \text{tolrel} \cdot \|f(x)\| + \text{tolabs}$
2. Do while  $\|f(x)\| > \text{tol}$ 
  1. Berechne Jacobi-Matrix  $F(x)$
  2. Zerlege  $F(x) = LU$
  3. Löse  $LU \cdot d = -f(x)$
  4.  $x = x + d$
  5. Berechne  $f(x)$
3. Return  $x$

von  $f(x_k)$  und die  $n^2$  partiellen Ableitungen von  $F(x_k)$  ebenfalls zu berechnen, was zu einem beträchtlichen Funktionsaufwand des Newton-Verfahrens führen kann. Der zugehörige Algorithmus 3 erfordert als Input die Funktionen  $f$  und  $F$ , einen Startwert  $x_0$  und die (absolute und relative) Toleranz  $\text{tolabs}, \text{tolrel}$ .

Wir wollen nun die Frage behandeln, unter welchen Bedingungen das Newton-Verfahren gegen eine Nullstelle  $x^*$  von  $f$  konvergiert. Entsprechend den angenommenen Vorausset-

zungen unterscheidet man drei Formen, die *lokale*, *semilokale* und *globale Konvergenz*:

- Wenn vorausgesetzt wird, dass eine Lösung  $x^*$  existiert und der Konvergenzsatz dann garantiert, dass eine (Kugel-)Umgebung  $K = K[x^*; \delta]$  dieses Punktes existiert, so dass das Iterationsverfahren für alle  $x_0 \in K$  gegen  $x^*$  konvergiert, so heißt dieses Verfahren *lokal konvergent*. Über die Lage und Größe der Umgebung  $K$  ist in praxi meist nichts bekannt; man weiß nur, dass für hinreichend nahe der Lösung liegende Startpunkte das Verfahren konvergieren wird.
- Muss die Existenz von  $x^*$  nicht vorausgesetzt werden und können die Voraussetzungen eines Konvergenzsatzes bereits am Startpunkt  $x_0$  verifiziert werden, so heißt das betreffende Verfahren *semilokal konvergent*. Dazu ist häufig eine Menge  $D$  zu beschreiben, so dass für Startpunkte  $x_0 \in D$  die Iteration konvergiert. Satz 9 liefert ein Beispiel für eine semilokale Konvergenzaussage.
- Ist  $D \subset \mathbb{R}^n$  ein im Allgemeinen großer *vorgegebener* Bereich (z. B. eine Kugelumgebung von 0, ein  $n$ -dimensionales endliches oder unendliches Intervall), so heißt ein Iterationsverfahren *global konvergent auf  $D$* , falls es für jeden Startpunkt  $x_0 \in D$  gegen eine Lösung  $x^*$  in  $D$  konvergiert. Verfahren dieser Art werden in Abschnitt 2.4 behandelt.

Wir wollen zuerst einen lokalen Konvergenzsatz für das Newton-Verfahren formulieren. Dazu betrachten wir Gleichung (21) und setzen die Existenz einer regulären Lösung  $x^* \in D$  voraus.

#### Definition 4 (Reguläre Lösung)

Eine Lösung  $x^* \in D$  heißt *regulär (isoliert)*, wenn

- eine Kugel  $K = K[x^*; \delta^*] = \{x \mid \|x - x^*\| \leq \delta^*\}$  um  $x^*$  mit  $K \in \text{int}(D)$  existiert,
- die Jacobi-Matrix  $F(x) = f'(x)$  auf  $K$  Lipschitz-stetig ist und
- die Jacobi-Matrix  $F(x^*)$  an der Lösung regulär ist.

Die Regularität der Nullstelle stellt in der Tat eine *Standardvoraussetzung* an das zu lösende Problem dar. Bei singulärer Jacobi-Matrix  $F(x^*)$  ist das Verhalten des Newton-Verfahrens überaus kompliziert: Während in einigen Fällen eine lineare Konvergenz eintritt (z. B. bei skalaren Nullstellenproblemen), versagt das Verfahren in höherdimensionalen Systemen häufig.

#### Bemerkung 5

1. Eine Lösung  $x^*$  ist *geometrisch isoliert*, falls eine Umgebung  $K$  existiert, in der keine weitere Lösung  $x^{**} \neq x^*$  liegt. Die Regularität einer Nullstelle darf deshalb nicht mit deren *geometrischer Isoliertheit* verwechselt werden. Aus der Regularität folgt im übrigen stets die geometrische Isoliertheit einer Lösung.

2. Falls  $f \in C^2(D)$  auf der gesamten Menge  $D$  ist, so lässt sich leicht zeigen, dass die Bedingung  $\text{rang } F(x^*) = n$  bzw.  $\det(F(x^*)) \neq 0$  hinreichend für die Regularität einer Lösung  $x^* \in \text{int}(D)$ . ■

### Beispiel 6

Für das System mit zweimal stetig differenzierbarem  $f \in \mathbb{R}^2$

$$f_1(x_1, x_2) = (x_1 - x_2)^2 = 0$$

$$f_2(x_1, x_2) = x_1 + x_2 - 2 = 0$$

erhält man die Jacobi-Matrix

$$F(x) = \begin{pmatrix} 2(x_1 - x_2) & -2(x_1 - x_2) \\ 1 & 1 \end{pmatrix}.$$

An der Lösung  $x^* = (1, 1)$  hat deren Determinante den Wert

$$\det(F(x^*)) = \begin{vmatrix} 0 & 0 \\ 1 & 1 \end{vmatrix} = 0,$$

weshalb  $x^*$  nicht regulär ist. Modifiziert man die Gleichungen zu

$$f_1(x_1, x_2) = (x_1 - x_2)^2 - 4 = 0$$

$$f_2(x_1, x_2) = x_1 + x_2 - 2 = 0,$$

so ändert man die Jacobi-Matrix  $F(x)$  damit nicht. Für die Lösung  $x^* = (2, 0)$  besitzt sie nun die Determinante

$$\det(F(x^*)) = \begin{vmatrix} 4 & -4 \\ 1 & 1 \end{vmatrix} = 8 \neq 0,$$

also ist diese Lösung regulär. ■

Vergleichen wir das Konvergenzverhalten des Newton-Verfahrens in diesem Beispiel miteinander, so lässt sich im regulären Fall eine wesentlich höhere Geschwindigkeit erkennen. Wir führen deshalb eine Konvergenzordnung ein und betrachten dazu eine Folge von Näherungen  $(x_k)$ , die mit einer allgemeinen Iteration  $x_{k+1} = g(x_k)$  erzeugt wird.

### Definition 7 (Q-Ordnung)

- i. Die Folge  $(x_k)$  konvergiert *mindestens mit der Q-Ordnung*  $p \geq 1$ , falls ein Konvergenzfaktor  $Q_+ > 0$  existiert, so dass für alle  $k \geq k_0$ ,  $k_0 \in \mathbb{N}$ ,

$$\|x_{k+1} - x^*\| \leq Q_+ \|x_k - x^*\|^p$$

gilt. Falls  $p = 1$  ist, so wird  $Q_+ < 1$  gefordert.

- ii. Die Folge  $(x_k)$  konvergiert *höchstens mit der Q-Ordnung*  $p \geq 1$ , falls ein Konvergenzfaktor  $Q_- > 0$  existiert, so dass für alle  $k \geq k_0$ ,  $k_0 \in \mathbb{N}$ , gilt:

$$\|x_{k+1} - x^*\| \geq Q_- \|x_k - x^*\|^p.$$

- iii. Die Folge  $(x_k)$  konvergiert *genau mit der Q-Ordnung*  $p \geq 1$ , falls sie mindestens und höchstens mit Q-Ordnung  $p$  konvergiert.
- iv. Das Iterationsverfahren  $x_{k+1} = g(x_k)$  konvergiert bezüglich  $x^*$  *mindestens mit der Q-Ordnung*  $p$ , wenn jede damit erzeugte Folge  $(x_k)$  mindestens mit Q-Ordnung  $p$  konvergiert.
- v. Existiert eine Folge  $(x_k)$  genau mit Q-Ordnung  $p$ , so konvergiert auch das Iterationsverfahren  $x_{k+1} = g(x_k)$  bezüglich  $x^*$  *genau mit Q-Ordnung*  $p$ .

Verfahren der Q-Ordnungen 1, 2 bzw. 3 werden als *linear*, *quadratisch* bzw. *kubisch konvergent* bezeichnet. Definition 7 setzt allerdings nicht voraus, dass die Fehler  $\|x_k - x^*\|$  verschieden von Null sind. Tritt dieser Fall für ein  $k = k_1 > k_0$  bei einem Verfahren ein, das mit *genau* der Q-Ordnung  $p$  konvergiert, so verschwinden alle Fehler der Iterationen für  $k \geq k_1$ . Andererseits bedeutet  $\|x_k - x^*\| > 0$  für ein  $k \geq k_0$ , dass alle weiteren Fehler ungleich Null sind. Wir setzen desweiteren diesen nichttrivialen Fall und ein Verfahren der genauen Q-Ordnung  $p$  voraus. Existiert der Grenzwert

$$q := \lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p},$$

so bezeichnet man ihn als *asymptotischen Konvergenzfaktor*. Falls jedoch  $q = 0$  ist, so erweist sich das Verfahren als *schneller als mit Q-Ordnung*  $p$  *konvergent*. Insbesondere heißen Iterationsverfahren mit

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0$$

*überlinear konvergent*.

Den beträchtlichen Vorteil eines Verfahrens mit Konvergenzordnung  $p > 1$  verdeutlicht man am besten, wenn man die Definition 7 rekursiv anwendet, was zur Darstellung

$$\|x_k - x^*\| \leq Q_+^{\frac{p^k - 1}{p - 1}} \|x_0 - x^*\|^p \quad (8)$$

führt. Bei Vorliegen linearer Konvergenz hätte sich dagegen

$$\|x_k - x^*\| \leq Q_+^k \cdot \|x_0 - x^*\| \quad (9)$$

ergeben. Angenommen, in beiden Fällen liege derselbe Anfangsfehler  $\|x_0 - x^*\| = 1.0$  vor und die Konvergenzkonstante sei  $Q_+ = 0.1$ . Dann schätzt man den Fehler des linear konvergenten Verfahrens durch

$$\|x_k - x^*\| \leq 10^{-k}$$



ab, d. h. pro Iterationsschritt gewinnt man durchschnittlich eine Dezimalziffer an Genauigkeit. Für ein quadratisch konvergentes Verfahren dagegen erhält man

$$\|x_k - x^*\| \leq 10^{-2^k},$$

so dass sich in jedem Iterationsschritt die Zahl der richtigen Ziffern verdoppelt!

Wir können nun einen lokalen Konvergenzsatz für das Newton-Verfahren angeben. Den umfangreichen Beweisgang dazu findet man z. B. bei Hoffmann et al. (2005).

### Satz 8 (Lokale Konvergenz)

$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  besitze die reguläre Lösung  $x^* \in D$ . Dann existieren Konstanten  $\delta > 0$  und  $Q > 0$  mit  $Q\delta < 1$  und eine Kugel  $K = K[x^*; \delta]$ , so dass folgende Behauptungen gelten:

- i. Das Newton-Verfahren ist für jeden Startwert  $x_0 \in K$  durchführbar mit  $x_k \in K$  für alle  $k \geq 0$ .
- ii.  $\lim_{k \rightarrow \infty} x_k = x^*$ .
- iii.  $x^*$  ist die einzige Lösung in der Kugel  $K$ .
- iv. Die Konvergenz ist mindestens Q-quadratisch, d. h.

$$\|x_{k+1} - x^*\| \leq Q\|x_k - x^*\|^2 \quad \forall k \geq k_0.$$

Da sich die Regularitätsvoraussetzung auf eine im allgemeinen unbekannte Lösung  $x^*$  bezieht, ist sie in der Praxis meist nicht überprüfbar. Geeigneter sind deshalb semilokale Sätze, die entsprechende Voraussetzungen an die vom Anwender des Verfahrens festgelegte *Startlösung*  $x_0$  fordern. Wir geben den bekanntesten Satz an; einen vollständigen Beweis findet man z. B. in Isaacson und Keller (1972).

### Satz 9 (L. V. Kantorovics)

Für die Startlösung  $x_0$  des Newton-Verfahrens seien folgende *Voraussetzungen* erfüllt:

- i. Es existiert die inverse Matrix  $F(x_0)^{-1}$  mit  $\|F(x_0)^{-1}\| \leq \alpha$ .
- ii. Für die Differenz der ersten beiden Näherungen gelte

$$\|x_1 - x_0\| = \|F(x_0)^{-1}f(x_0)\| \leq \beta.$$

- iii.  $f$  sei 2-mal stetig differenzierbar mit

$$\sum_{k=1}^n \left| \frac{\partial^2 f_i(x)}{\partial x_j \partial x_k} \right| \leq \frac{\gamma}{n}, \quad i, j = 1(1)n,$$

für alle  $x \in K = K[x_0; 2\beta] = \{x \mid \|x - x_0\| \leq 2\beta\}$ .

- iv.  $\alpha\beta\gamma \leq 1/2$ .

Dann gelten folgende *Behauptungen*:

- i. Das Newton-Verfahren ist für den Startwert  $x_0 \in K$  unbeschränkt *durchführbar* mit  $x_k \in K$  für alle  $k \geq 0$ .
- ii. Die Folge  $(x_k)$  konvergiert gegen ein  $x^*$  mit  $f(x^*) = 0$ .
- iii. Für den Fehler gilt

$$\|x_k - x^*\| \leq \frac{2\beta}{2^k} \quad k = 0, 1, 2, \dots$$

Falls in Voraussetzung iv. sogar  $\alpha\beta\gamma < 1/2$  gilt, so kann zusätzlich nachgewiesen werden, dass die Folge  $(x_k)$  mindestens Q-quadratisch konvergiert.

## 2.2 Vereinfachte Newton-Verfahren

Für großdimensionale Systeme ist der Aufwand pro Iterationsschritt auch in der praktikablen Form 7 des Newton-Verfahrens zu hoch. Man approximiert die Jacobi-Matrix  $F(x_k)$  dann durch eine geeignete konstante Matrix  $A$ . Im einfachsten Fall kann man  $A = F(x_0)$  wählen und erhält so das *vereinfachte Newton-Verfahren* (*Sehnenverfahren*, engl. *chord method*):

- (1) Berechne die Matrix  $A = F(x_0)$  und deren LU-Zerlegung.
- (2) Löse für  $k = 0, 1, 2, \dots$  das lineare Gleichungssystem

$$A d^k = -f(x_k), \quad x_{k+1} = x_k + d^k. \quad (10)$$

Offenbar ist nun jeder Iterationsschritt wesentlich effizienter, denn der arithmetische Aufwand für eine Vorwärts-Rückwärts-Elimination ist nur von der Ordnung  $\mathcal{O}(n^2)$ , und es sind lediglich  $n$  Funktionswerte in  $f(x_k)$  zu berechnen. Dem steht eine langsamere Konvergenz gegenüber. Im Gegensatz zum Newton-Verfahren kann für das Sehnenverfahren nur lineare Konvergenz erwartet werden. Um die Konvergenz des Verfahrens zu garantieren, wird wie im Falle des Newton-Verfahrens die Regularität der Nullstelle  $x^*$  vorausgesetzt (vgl. Kelley (1995)).

### Satz 11 (Lokale Konvergenz)

$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  besitze die reguläre Nullstelle  $x^* \in D$ . Dann existieren Konstanten  $\delta > 0$  und  $Q_1 > 0$ , so dass folgende Behauptungen gelten:

- i. Das Sehnenverfahren ist für jeden Startwert  $x_0 \in K = K[x^*; \delta]$  durchführbar mit  $x_k \in K$  für alle  $k \geq 0$ .
- ii.  $\lim_{k \rightarrow \infty} x_k = x^*$ .

**ALGORITHMUS 10 (Sehnenverfahren)**

Function  $[x] = \text{newton0}(f, F, x_0, \text{tolabs}, \text{tolrel})$

1. Berechne  $x = x_0$ ,  $\text{tol} = \text{tolrel} \cdot \|f(x)\| + \text{tolabs}$
2. Berechne Jacobi-Matrix  $F(x)$
3. Zerlege  $F(x) = LU$
4. Do while  $\|f(x)\| > \text{tol}$ 
  1. Löse  $LU \cdot d = -f(x)$
  2.  $x = x + d$
  3. Berechne  $f(x)$
5. Return  $x$

iii. Die Konvergenz ist Q-linear mit der Fehlerschätzung

$$\|x_{k+1} - x^*\| \leq Q_1 \|x_0 - x^*\| \|x_k - x^*\|, \quad k = 0, 1, 2, \dots \quad (11)$$

Algorithmus 10 erfordert als Input die Funktionen  $f$  und  $F$ , einen Startwert  $x_0$  und die (absolute und relative) Toleranz  $\text{tolabs}, \text{tolrel}$ . Auch diesen Algorithmus kann man leicht als MATLAB-Funktion implementieren.

Falls sich die Jacobi-Matrix  $F(x_k)$  von Iteration zu Iteration nur wenig ändert, empfehlen sich mehrere Zwischenschritte mit ein- und derselben Matrix  $A$ . Dafür ist jedesmal nur eine LU-Zerlegung erforderlich. Man erhält so eine Kombination aus Newton- und Sehnenverfahren, die häufig als *Shamanskii-Verfahren* bezeichnet wird. Es lässt sich als eine Variante des Newton-Verfahrens mit  $m$  inneren Iterationsschritten interpretieren:

Für  $k = 0, 1, 2, \dots$  iteriere

$$\begin{aligned} y_0 &= x_k - [F(x_k)]^{-1} f(x_k), \\ y_{j+1} &= y_j - [F(x_k)]^{-1} f(y_j), \quad j = 0(1)m - 1 \\ x_{k+1} &= y_m. \end{aligned} \quad (12)$$

Das Newton-Verfahren liegt offenbar im Sonderfall  $m = 0$  vor, während im Grenzfall  $m \rightarrow \infty$  dieses Verfahren in das Sehnenverfahren übergeht.

Nimmt man die Werte  $y_j$  der inneren Iteration als Hilfsgrößen an und betrachtet lediglich die Folge  $(x_k)$  der äußeren Iterierten, so kann man unter geeigneten Voraussetzungen nachweisen, dass diese Folge mindestens mit Q-Ordnung  $m + 2$  gegen die Lösung  $x^*$  konvergiert. So kann mit relativ geringem Zusatzaufwand (eine Berechnung von  $f$  und eine Vorwärts-Rückwärts-Elimination) eine höhere Konvergenzordnung als beim gewöhnlichen Newton-Verfahren erreicht werden.

Der Algorithmus 12 erfordert als Input ebenfalls die Funktionen  $f$  und  $F$ , einen Startwert  $x_0$  und die (absolute und relative) Toleranz  $tolabs, tolrel$ . Hinzu kommt die Anzahl  $m$  der inneren Iterationen. Sämtliche Iterationswerte können in ein- und derselben Variablen  $x$

### ALGORITHMUS 12 (Shamanskii-Verfahren)

Function  $[x] = \text{shamanskii}(f, F, x_0, tolabs, tolrel, m)$

1. Berechne  $x = x_0$ ,  $tol = tolrel \cdot \|f(x)\| + tolabs$
2. Do while  $\|f(x)\| > tol$ 
  1. Berechne Jacobi-Matrix  $F(x)$
  2. Zerlege  $F(x) = LU$
  3. For  $j = 0(1)m$ 
    1. Löse  $LU \cdot d = -f(x)$
    2.  $x = x + d$
    3. Berechne  $f(x)$
    4. If  $\|f(x)\| < tol$  Return  $x$
3. Return  $x$

gespeichert werden. Wird die geforderte Genauigkeit bereits in einer inneren Iteration erreicht, so kann der entsprechende Iterationswert zurückgegeben werden.

### Beispiel 13

Wir betrachten die Gleichungen

$$f_1(x_1, x_2) = 1 - x_1^2 - x_2^2 = 0, \quad f_2(x_1, x_2) = -2x_1 + x_2 = 0$$

und wenden (a) das Newton-Verfahren, (b) das Shamanskii-Verfahren mit den inneren Iterationszahlen  $m = 1, 2, 5, 10$  und (c) das Sehnungsverfahren mit denselben Parametersätzen an. In Tabelle 3 wird die Anzahl  $k$  aller Iterationen des Newton-, des Shamanskii- und des Sehnungsverfahrens gegenübergestellt. In den mit \* markierten Fällen wurde die geforderte Genauigkeit  $tol$  nicht vollständig erreicht. Beachtet man jedoch, dass das Newton-Verfahren in dritten Aufruf  $k = 6$  Berechnungen der Jacobi-Matrix und einen arithmetischen Aufwand von  $k \cdot \frac{2}{3}n^3$  erfordert, so liefert das Shamanskii-Verfahren die Lösung bereits für  $k = 16$ ,  $m = 10$  mit  $\lfloor \frac{k-1}{m+1} \rfloor + 1 = 2$  Jacobi-Matrizen und einem arithmetischen Aufwand von  $2 \cdot \frac{2}{3}n^3$ . ■

Der Vorteil dieses Verfahrens tritt besonders bei großdimensionalen Gleichungssystemen stärker hervor. Es kann zudem verbessert werden, wenn die Anzahl  $m$  ergebnisabhängig gewählt und gegebenenfalls adaptiv geändert wird. Das kann durch Kontrolle der Fehlerentwicklung der inneren Iteration erfolgen (vgl. Kelley (1995)).

**Tab. 1** Anzahl der benötigten Iterationen zu Beispiel 13.

tolabs tolrel	Startpunkt ( $x_1, x_2$ )	Newton- Verfahren	Shamanskii-Verfahren				Sehnen- verfahren
			$m = 1$	$m = 2$	$m = 5$	$m = 10$	
$10^{-6}$	(3,1.7)	5	7	7	10	11*	28*
$10^{-12}$	(3,1.7)	6	8	9	13	16	50*
$10^{-12}$	(-3,-2)	6	8	10	13	16	50*

## 2.3 Newton-Verfahren mit Differenzenquotienten

Häufig sind die Ableitungen von  $f$  nur mit erheblichem Aufwand exakt zu berechnen. Dann kann man das für skalare Gleichungen bekannte Sekantenverfahren verallgemeinern und approximiert die Jacobi-Matrix  $F(x_k) = f'(x_k)$  durch eine leichter zu berechnende Matrix  $A_k \in \mathbb{R}^{n \times n}$  mittels Differenzenquotienten

$$x_{k+1} = x_k - A_k^{-1} f(x_k), \quad k = 0, 1, 2, \dots \quad (13)$$

und einem geeigneten Schrittweitenvektor  $h = (h_1, \dots, h_n)^T$ ,  $h_j > 0$ . Derartige Verfahren bezeichnet man oft als *Newton-ähnlich* (*Newton-like*), wenn die erzeugten Folgen  $(x_k)$ ,  $k = 0, 1, 2, \dots$ , das Verhalten der Newton-Folge aus Verfahren 3 approximieren. Besitzt die Gleichung  $f(x) = 0$  die reguläre Lösung  $x^* \in D \subset \mathbb{R}^n$ , so definiert man genauer:

### Definition 14 (Newton-ähnliches Verfahren)

Eine konvergente Näherungsfolge  $(x_k)$ ,  $k = 0, 1, 2, \dots$ , mit  $x_k \neq x^*$  in einer Umgebung  $S$  von  $x^*$  heißt *Newton-ähnlich*, wenn

$$\lim_{k \rightarrow \infty} \frac{\|f(x_k) + f'(x_k)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0$$

gilt. Ist jede Folge mit  $x_0 \in S$  Newton-ähnlich, so bezeichnet man das Verfahren selbst als *Newton-ähnlich*.

Für das Newton-Verfahren selbst trifft die Definition offenbar zu. Man kann unter der Voraussetzung einer regulären Lösung  $x^*$  nachweisen, dass ein Verfahren genau dann überlinear konvergiert, wenn es Newton-ähnlich ist (vgl. Schwetlick (1979)).

Betrachten wir nun Verfahren der Gestalt (13) und stellen die Matrix  $A_k$  durch Differenzenquotienten dar. Als einfachste Approximation  $A_k$  bietet sich die Einpunkt-Approximation mittels der Vorwärts-Differenzenquotienten von  $f'(x)$

$$\{\nabla f(x, h)\}_{ij} = \frac{1}{h_j} [f_i(x_1, \dots, x_j + h_j, \dots, x_n) - f_i(x)], \quad i, j = 1 \dots n,$$

an, die mit dem  $j$ -ten Einheitsvektor  $e_j$  spaltenweise als

$$\nabla f(x, h)e_j = \frac{1}{h_j} [f(x + h_j e_j) - f(x)], \quad j = 1 \dots n \quad (14)$$

notiert werden kann. Zuerst ist zu klären, unter welchen Bedingungen das entstehende Verfahren (13) durchführbar ist und die erzeugte Folge der Näherungslösungen  $(x_k)$ ,  $k = 0, 1, 2, \dots$  gegen eine Lösung  $x^*$  konvergiert. Da anstelle der einfachen Differenzenquotienten (14) auch andere Approximationen denkbar sind, soll folgende Verallgemeinerung eingeführt werden:

**Definition 15 (Streng konsistente Approximation)**

$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  sei auf  $D_0 \subset \text{int } D$  differenzierbar. Die Schrittweitenmenge sei

$$H = \{h = (h_1, h_2, \dots, h_n) \in \mathbb{R}^n \mid h_i \neq 0, i = 1, \dots, n, \|h\| \leq r\}.$$

Dann heißt die Abbildung  $\nabla f : D_0 \times H \rightarrow \mathbb{R}^{n \times n}$  *streng konsistente Approximation* für  $f'(x)$  auf  $D_0$  mit dem Diskretisierungsbereich  $H \subset \mathbb{R}^n$ , wenn  $0 \in \mathbb{R}^n$  Häufungspunkt von  $H$  ist und eine Konstante  $C > 0$  existiert, so dass

$$\|\nabla f(x, h) - f'(x)\| \leq C \cdot \|h\| \quad (15)$$

für alle  $(x, h) \in D_0 \times H$  gilt.

Mit der Voraussetzung  $f \in C^2(\mathbb{R}^n)$  erfüllt unsere Differenzenapproximation (14) offenbar die Bedingung (15), d. h.  $\lim_{h \rightarrow 0} \nabla f(x, h) = f'(x)$  mit Ordnung 1 in  $h$ . Bei Schwetlick (1979) wird dann die Konvergenz des Verfahrens (13) mit folgendem Satz nachgewiesen:

**Satz 16 (Konvergenz)**

$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  besitze die reguläre Lösung  $x^* \in D$ , und  $\nabla f : D_0 \times H \rightarrow \mathbb{R}^{n \times n}$  sei eine streng konsistente Approximation von  $f'$  auf einer Umgebung  $D_0$  von  $x^*$ . Dann existieren Konstanten  $\delta_0 > 0$  und  $\varepsilon > 0$ , so dass folgende Behauptungen gelten:

- i. Das Verfahren (13) ist für jedes  $x_0 \in K_0 := K[x^*; \delta_0]$  durchführbar, und  $x_k \in K_0$  für  $k = 0, 1, 2, \dots$ , wenn in jedem Schritt  $h_k \in H$  so gewählt wird, dass  $\|h_k\| < \varepsilon$  garantiert ist.
- ii. Die Folge  $(x_k)$  konvergiert mindestens linear gegen  $x^*$ .
- iii. Falls zusätzlich  $\lim_{k \rightarrow \infty} h_k = 0$  gilt, so ist die Konvergenz der  $x_k$  überlinear.

Damit ist das Verfahren Newton-ähnlich im Sinne der angegebenen Definition, falls die Schrittweitenfolge  $(h_k)$  gegen Null konvergiert. In praxi können allerdings wegen der begrenzten Stellenzahl der benutzten Gleitpunktzahlen die Schrittweiten  $h_j$  der Differenzenapproximation (14) nicht beliebig nahe bei Null gewählt werden, sondern müssen

auf jeden Fall größer als die relative Maschinengenauigkeit  $\varepsilon_M$  sein. Damit erfüllt die reale Schrittweitenmenge  $H_{real}$  nicht die Voraussetzung aus Definition 15 und der Konvergenzsatz ist nur bedingt aussagekräftig. Beantworten wir deshalb die Frage, wie die

### ALGORITHMUS 17 (Newton-Verfahren mit Differenzen)

Function  $[x] = \text{newtondiff}(f, x_0, \text{tolabs}, \text{tolrel})$

1. Berechne  $x = x_0$ ,  $\text{tol} = \text{tolrel} \cdot \|f(x)\| + \text{tolabs}$
2. Do while  $\|f(x)\| > \text{tol}$ 
  1. Wähle  $h_j = \sqrt{\varepsilon_M} (1 + |x_j|)$ ,  $j = 1(1)n$
  2. Approximiere die Jacobi-Matrix durch  $A_k$  mit
$$A_k e_j = \frac{1}{h_j} [f(x + h_j e_j) - f(x)], \quad j = 1(1)n$$
  3. Löse  $A_k \cdot d = -f(x)$  nach  $d$
  4.  $x = x + d$
  5. Berechne  $f(x)$
3. Return  $x$

Diskretisierungsschrittweiten  $h_j$  in  $h = (h_1, h_2, \dots, h_n)$  passend zu wählen sind. Neben dem Diskretisierungsfehler  $f'(x) - \nabla f(x, h)$  sind dazu nun auch andere Fehlerquellen, insbesondere Rundungsfehler, zu berücksichtigen. Wird nämlich anstelle von  $f(x)$  der fehlerbehaftete Funktionswert  $\tilde{f}(x) = f(x) + \varepsilon(x)$ ,  $\|\varepsilon(x)\| < \varepsilon_M$  mit der Maschinengenauigkeit  $\varepsilon_M$  benutzt, so erhält man mit der Voraussetzung  $f \in C^2(\mathbb{R}^n)$  für den gesamten Approximationsfehler der Differenzenapproximation die Abschätzung

$$\begin{aligned} \|f'(x)e_j - \nabla \tilde{f}(x, h)e_j\| &= \|f'(x)e_j - \frac{1}{h_j} [\tilde{f}(x + h_j e_j) - \tilde{f}(x)]\| \\ &\leq \|f'(x)e_j - \frac{1}{h_j} [f(x + h_j e_j) - f(x)]\| + \frac{2 \max \|\varepsilon(x)\|}{h_j} \\ &\leq Ch_j + \frac{2\varepsilon_M}{h_j} \end{aligned}$$

mit einer funktionsabhängigen Konstanten  $C > 0$ . Minimierung der rechten Seite über  $h_j$  ergibt den Minimalwert  $h_j = \sqrt{2\varepsilon_M/C}$ , womit  $h = \mathcal{O}(\sqrt{\varepsilon_M})$  gilt. Um für betragsgroße Werte  $x_j$  auch entsprechend große Diskretisierungsschrittweiten  $h_j$  zu erhalten, hat sich eine kombinierte Absolut-Relativ-Wahl

$$h_j = \sqrt{\varepsilon_M} (1 + |x_j|), \quad j = 1(1)n \quad (16)$$

gut bewährt. Damit wird zugleich garantiert, dass für Werte von  $x_j$  nahe Null keine Stellenauslöschung in den Differenzenquotienten erfolgt. Das entstehende Verfahren ist allerdings nun nicht mehr Newton-ähnlich im engeren Sinne der Definition.

## 2.4 Gedämpfte Newton-Verfahren und globale Konvergenz

Die bisher behandelten Verfahren sind lokal konvergent, da unter geeigneten Voraussetzungen (Glattheit, Regularität) stets eine Konvergenzumgebung  $S$  der Nullstelle  $x^*$  der Gleichung

$$f(x) = 0, \quad f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad D \text{ offen} \quad (17)$$

existiert. Über die Lage und Größe der Umgebung  $S$  ist in praxi meist nichts bekannt. Um den semilokalen Konvergenzsatz 9 von L.V. Kantorovics anwenden zu können, sind schwer nachzuvollziehende Abschätzungen in der Umgebung des Startwertes  $x_0$  durchzuführen, der zudem nahe der Lösung  $x^*$  liegen sollte.

Die Menge  $S(x^*)$  aller Startwerte  $x_0$ , für die das jeweils betrachtete Verfahren gegen die Nullstelle  $x^*$  konvergiert, bildet den *Einzugsbereich von  $x^*$* . Um ihn zu vergrößern, kann man zur Bestimmung der Nullstelle  $x^*$  von  $f$  auch das zugehörige Minimierungsproblem

$$\psi(x) = \frac{1}{2} \|f(x)\|^2 \Rightarrow \text{Min!} \quad \text{über } x \in D \subset \mathbb{R}^n, \quad (18)$$

mit der Euklidischen Norm  $\|x\|$  betrachten. Gesucht ist dann eine Minimalstelle  $x^*$  mit  $\psi(x^*) = 0$ . Ausgehend von einer  $k$ -ten Lösungsnäherung  $x_k$  lässt sich der neue Wert  $x_{k+1} = x_k + d_k$  mit der Newton-Richtung (bei Minimierungsverfahren auch „Such-Richtung“ genannt)  $d_k = -f'(x_k)^{-1} f(x_k)$  ermitteln und anschließend kontrollieren, ob die *einfache Abstiegsbedingung*

$$\psi(x_{k+1}) < \psi(x_k)$$

für die Funktion  $\psi$  erfüllt ist. Andernfalls kann man den Korrekturvektor  $d_k$  mit einem Faktor  $\tau_k \in (0, 1]$  dämpfen und mit dem zurückgesetzten Testpunkt

$$x_{k+1} = x_k + \tau_k \cdot d_k \quad \text{mit} \quad d_k = -f'(x_k)^{-1} f(x_k) \quad (19)$$

die Abstiegsbedingung erneut überprüfen. Nach eventuell weiteren erforderlichen Reduzierungen  $\tau_k := \tau_k \cdot \alpha$  mit dem fest gewählten Skalierungsfaktor  $\alpha \in (0, 1)$  liefert der erste akzeptierte  $\tau_k$ -Wert den *Dämpfungsfaktor* des sogenannten *gedämpften Newton-Verfahrens* (19). Häufig wird eine sukzessive Halbierung mit  $\alpha = 0.5$  vorgenommen. Bei Minimierungsverfahren hat es sich bewährt, diese einfache Abstiegsbedingung durch eine Bedingung für *hinreichenden Abstieg*

$$\psi(x_{k+1}) < (1 - 2\delta\tau_k) \psi(x_k), \quad \text{mit einer Konstanten } \delta \in (0, \frac{1}{2}) \quad (20)$$

zu ersetzen. Offenbar ist  $1 - 2\delta\tau_k \in (0, 1)$ , wobei Werte  $\delta = 10^{-2}, 10^{-4}$  üblich sind. Der erste  $\tau_k$ -Wert, mit dem diese Bedingung erfüllt ist, liefert ein gedämpftes Newton-Verfahren (19) mit der sogenannten *Armijo-Regel* (20), die in Algorithmus 18 implementiert wurde. Eine detaillierte Begründung dieser und weiterer verbesserter Strategien findet man bei Kosmol (1993).



Das Hauptziel, den Einzugsbereich von  $x^*$  zu vergrößern, kann mit dem gedämpften Newton-Verfahren oft erreicht werden. Ist  $S \subset \mathbb{R}^n$  ein im Allgemeinen großer vorgegebener Bereich, so ist das Iterationsverfahren *global konvergent auf  $S$* , falls es für jeden Startpunkt  $x_0 \in S$  gegen einen Fixpunkt  $x^*$  in  $S$  konvergiert. Kann die Regularität der Jacobi-Matrix  $f'(x)$  auf einer ganzen Niveaumenge vorausgesetzt werden, so lässt sich die globale Konvergenz des gedämpften Newton-Verfahrens für alle Startwerte aus dieser Menge nachweisen (vgl. Schwetlick (1979) zum aufwändigen Beweis des Satzes):

### Satz 19 (Globale Konvergenz)

Es sei  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$  2-mal stetig differenzierbar auf der offenen Menge  $D$ . Zu einem festem  $y \in D$  werde die Niveaumenge

$$S = \{x \in D \mid \|f(x)\| \leq \|f(y)\|\}$$

definiert, die als kompakt vorausgesetzt wird. Ist die Jacobi-Matrix  $f'(x)$  für alle  $x \in S$  invertierbar, so gelten folgende Behauptungen:

- i. Das gedämpfte Newton-Verfahren (19) mit Armijo-Regel (20) ist für jedes  $x_0 \in S$  durchführbar mit  $x_k \in S$  und streng monoton fallenden Funktionswerten  $\|f(x_k)\|$ .
- ii. Die Folge  $(x_k)$  konvergiert gegen eine Nullstelle  $x^*$  von  $f$  in der Menge  $S$ .
- iii. Falls das Verfahren nicht nach endlich vielen Schritten mit einer Nullstelle endet, so existiert ein  $K \in \mathbb{N}$ , bei dem das Verfahren für  $k > K$  in das gewöhnliche Newton-Verfahren übergeht und Q-quadratische Konvergenz eintritt.

### ALGORITHMUS 18 (Gedämpftes Newton-Verfahren)

Function  $[x, y] = \text{newtonarmijo}(f, F, x_0, \text{tolabs}, \text{tolrel})$

1. Berechne  $x = x_0$ ,  $\text{tol} := \text{tolrel} \cdot \|f(x)\| + \text{tolabs}$
2. Wähle Konstanten  $\delta \in (0, \frac{1}{2})$ ,  $\alpha \in (0, 1)$  und  $\lambda := 1$
3. Do while  $\|f(x)\| > \text{tol}$ 
  1. Wähle  $\lambda := \min(\lambda/\alpha, 1)$
  2. Berechne Jacobi-Matrix  $F(x)$
  3. Berechne Newton-Richtung  $d$  aus  $F(x) \cdot d = -f(x)$
  4. Berechne Testpunkt  $y := x + \lambda d$
  5. Do while  $\|f(y)\|^2 > (1 - 2\delta\lambda)\|f(x)\|^2$ 
    1. Reduziere  $\lambda := \lambda \cdot \alpha$
    2. Berechne Testpunkt  $y := x + \lambda d$
  6. Aktualisiere  $x := y$
4. Return  $x, y = f(x)$

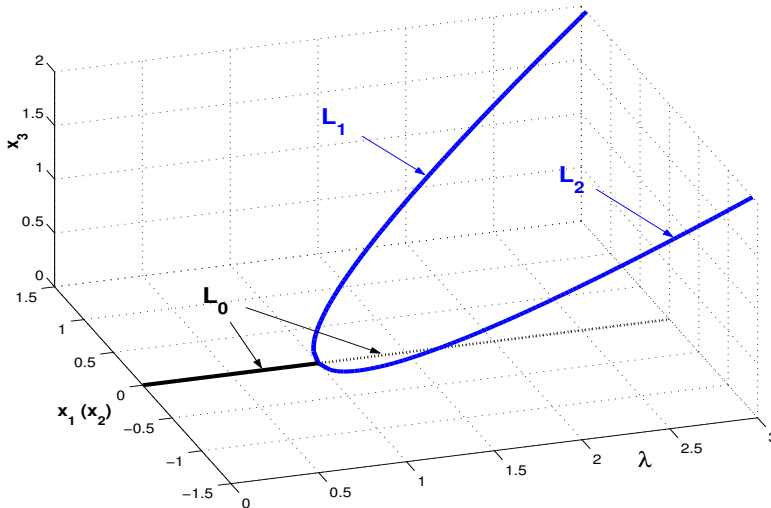
Die hier eingeführte Dämpfungsstrategie lässt sich nunmehr auch erfolgreich auf das vereinfachte Newton-Verfahren, das Shamanskii-Verfahren und das Newton-Verfahren mit Differenzenquotienten anwenden und dafür globale Konvergenz erreichen. Die komplizierte Konvergenztheorie übersteigt jedoch den Rahmen dieses Beitrags. Sehr gute und detaillierte Darstellungen des Stoffes findet man bei Deuffhard (2004) und Kelley (1995).

### 3 Parametrisierung von Lösungskurven und Fortsetzungsmethoden

Wir kehren nun zum eingangs genannten Problem (1) zurück und bestimmen parameterabhängige Gleichgewichtspunkte stets als Lösungen des Gleichungssystems

$$f(x, \lambda) = 0, \quad f : D \times \Lambda \subset \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n \quad (21)$$

mit skalarem Parameter  $\lambda \in \Lambda = [a, b]$ .



**Abb. 1**  
Lösungszweige  $L_0$  (schwarz) und  $L_1, L_2$  (blau) im Lösungsdiagramm zu Beispiel 20

#### Beispiel 20

Für  $\lambda \in [0, 3]$  berechnen wir alle Gleichgewichtspunkte des Lorenz-Systems (3) mittels

$$\begin{aligned} x_2 - x_1 &= 0 \\ \lambda x_1 - x_2 - x_1 x_3 &= 0 \\ x_1 x_2 - x_3 &= 0. \end{aligned}$$

Im Intervall  $0 \leq \lambda \leq 1$  existiert lediglich die asymptotisch stabile triviale Lösung  $L_0 : x_1 = x_2 = x_3 = 0$ , während für  $1 < \lambda \leq 3$  außer dieser Lösung  $L_0$  die beiden nichttrivialen Lösungen

$$L_1 : x_1 = x_2 = \sqrt{\lambda - 1}, x_3 = \lambda - 1 \quad \text{und} \quad L_2 : x_1 = x_2 = -\sqrt{\lambda - 1}, x_3 = \lambda - 1$$

gefunden werden. Wegen  $x_1 = x_2$  für alle Lösungen wählen wir eine dieser beiden Koordinaten aus und stellen die Lösungsbranche im  $(\lambda, x_1, x_3)$ -Lösungsdiagramm der Abbildung 1 dar. Beim Durchlaufen des Parameterwertes  $\lambda^* = 1$  tritt offenbar eine Lösungsverzweigung ein. Zugleich wird für  $\lambda > \lambda^*$  die triviale Lösung  $L_0$  instabil. ■

Abbildung 1 legt nahe, die Variablen  $x$  und  $\lambda$  zu einer einzigen Variablen  $y = (x, \lambda)$  mit  $y \in \mathbb{R}^{n+1}$  zusammenzufassen und System (21) als unterbestimmtes Gleichungssystem

$$f(y) = 0, \quad f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n, \quad f \in C^r(\mathbb{R}^{n+1}) \quad \text{mit} \quad r \geq 1, \quad (22)$$

in  $\mathbb{R}^{n+1}$  zu behandeln. Um dessen Lösungsmenge  $\mathcal{M}$  numerisch bestimmbar zu machen, treffen wir eine Zusatzannahme mit folgender

**Definition 21 (Regulärer Punkt, regulärer Wert)**

- i. Ein Punkt  $y \in \mathbb{R}^{n+1}$  heißt *regulärer Punkt* der glatten Abbildung  $f$ , wenn die Jacobi-Matrix  $Df(y) \in \mathbb{R}^{(n+1) \times n}$  vollen Rang hat, d. h.  $\text{rang}(Df(y)) = n$ . Andernfalls heißt  $y$  *singulärer Punkt*.
- ii. Ein Bildpunkt  $w \in \mathbb{R}^n$  heißt *regulärer Wert* von  $f$ , wenn die Menge

$$f^{-1}(w) := \{y \in D \times \Lambda \subset \mathbb{R}^{n+1} \mid f(y) = w\}$$

nur aus regulären Punkten von  $f$  besteht. Andernfalls heißt  $w$  *singulärer Wert*.

Wir nehmen desweiteren an, dass  $0 \in \mathbb{R}^n$  ein regulärer Wert von  $f$  ist, also die Jacobi-Matrix  $Df(y)$  vollen Rang für alle Lösungen von  $f(y) = 0$  besitzt. Mit

$$\mathcal{M} = \{y \in \mathbb{R}^{n+1} \mid f(y) = 0, \quad \text{rang}(Df(y)) = n\} \quad (23)$$

bezeichnen wir die Menge aller Lösungen von (22), die reguläre Punkte von  $f$  sind.

**Beispiel 22**

Mit  $y = (x_1, x_2, x_3, \lambda) = (y_1, y_2, y_3, y_4)$  lautet die Jacobi-Matrix des Lorenz-Systems

$$Df(y) = \begin{pmatrix} -1 & 1 & 0 & 0 \\ y_4 - y_3 & -1 & -y_1 & y_1 \\ y_2 & y_1 & -1 & 0 \end{pmatrix} \quad \text{wegen} \quad f(y) = \begin{pmatrix} y_2 - y_1 \\ y_4 y_1 - y_2 - y_1 y_3 \\ y_1 y_2 - y_3 \end{pmatrix},$$

womit sich für den Lösungsweig  $L_0 : y_0 = (0, 0, 0, y_4)$  nur dann  $\text{rang} Df(y_0) = 3$  ergibt, wenn  $y_4 \equiv \lambda \neq 1$  ist. Der Punkt  $y_0 = (0, 0, 0, 1)$  ist hingegen singulär. Also ist  $0 \in \mathbb{R}^3$  kein regulärer Wert von  $f$ . Schränken wir den Parameterbereich jedoch auf  $\Lambda = [1.5, 3]$  ein, so sind alle entsprechenden Lösungen reguläre Punkte von  $f$ . ■

### 3.1 Natürliche Parametrisierung

Ist für einen Lösungspunkt  $y_0 = (x_0, \lambda_0)$  die Teilmatrix  $D_x f(x_0, \lambda_0) \in \mathbb{R}^{n \times n}$  regulär, so kann der natürliche Parameter  $\lambda$  zur lokalen Parametrisierung der Lösungsmenge  $\mathcal{M}$  benutzt werden. Nach dem impliziten Funktionentheorem (vgl. Hoffmann et al. (2005)) existiert lokal eine Funktion  $x : [\lambda_0 - \varepsilon, \lambda_0 + \varepsilon] \rightarrow \mathbb{R}^n$  mit  $\varepsilon > 0$  und  $(x(\lambda), \lambda) \in \mathcal{M}$ . Ist  $D_x f(x, \lambda)$  sogar regulär für alle  $(x, \lambda) \in D \times \Lambda$ , so ist die Funktion  $x(\lambda)$  auf ganz  $\Lambda$  definiert. Wir treffen dazu folgende

**Voraussetzung 23 (Regularität)**

Es sei  $D \subset \mathbb{R}^n$  offen und  $\Lambda = [a, b]$ .  $f \in C^r(D \times \Lambda)$ ,  $r \geq 2$ , und  $D_x f(x, \lambda)$  ist regulär für alle  $(x, \lambda) \in D \times \Lambda$ .

Mit dieser Regularitätsvoraussetzung gewinnen wir den für das Verfahren grundlegenden

**Satz 24**

Unter Voraussetzung 23 existiert die implizit definierte Funktion  $x : \Lambda \rightarrow D$  mit den Eigenschaften:

- i.  $f(x, \lambda) = 0$  besitzt für jedes  $\lambda \in \Lambda$  genau eine Lösung  $x = x(\lambda) \in D$ .
- ii. Es gilt die Abschätzung  $\|x - x(\lambda)\| \leq M_0 \cdot \|f(x, \lambda)\| \quad \forall (x, \lambda) \in D \times \Lambda$ ,  $M_0 > 0$ .
- iii.  $x(\lambda)$  ist in  $\lambda \in \Lambda$  stetig differenzierbar und genügt der Einbettungsgleichung

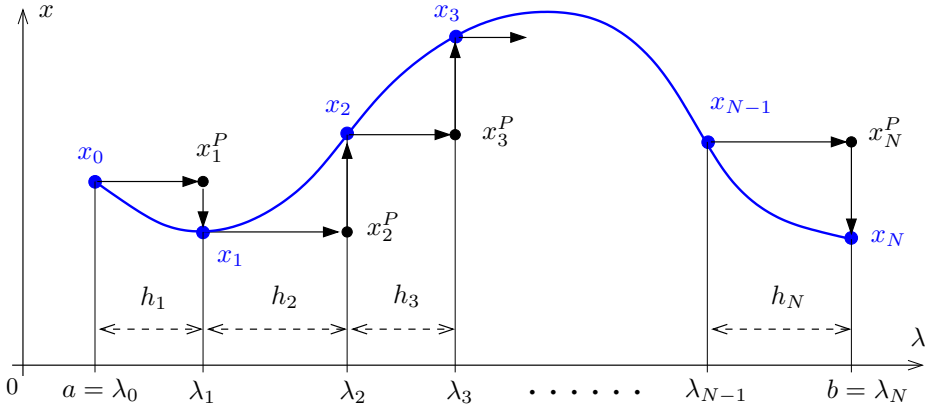
$$x'(\lambda) = -[D_x f(x(\lambda), \lambda)]^{-1} D_\lambda f(x(\lambda), \lambda). \quad (24)$$

Von einer direkten numerischen Integration dieser Einbettungsgleichung – mitunter auch als *Davidenko-DGL* bezeichnet – als Anfangswertproblem sollte wegen der komplizierten rechten Seiten und evtl. fehlender Lösungsstabilität Abstand genommen werden. Numerische Fortsetzungsmethoden nutzen vielmehr die Bestimmungsgleichungen (21) zur genauen Berechnung von  $x(\lambda)$ . Dazu unterteilen wir wie in Abbildung 2 das Parameterintervall  $\Lambda = [a, b]$  in  $N$  Teilintervalle mit den Teilpunkten  $a = \lambda_0 < \lambda_1 < \dots < \lambda_N = b$  und bestimmen sukzessive die Punkte  $(x(\lambda_j), \lambda_j)$  der Lösungskurve

$$\mathcal{L} = \{(x, \lambda) | (x, \lambda) \in D \times \Lambda \text{ mit } f(x, \lambda) = 0\}, \quad (25)$$

beginnend mit  $(x(a), a) = (x_0, \lambda_0)$  in der Reihenfolge  $(x_0, \lambda_0), (x_1, \lambda_1), \dots, (x_N, \lambda_N)$ . Dabei soll abkürzend  $x_j = x(\lambda_j)$  notiert werden. Ein *Fortsetzungsschritt* von  $\lambda_{j-1}$  bis  $\lambda_j$  ist dann folgendermaßen aufgebaut:

1. Gegeben sei ein (zuvor ermittelter) Kurvenpunkt  $(x_{j-1}, \lambda_{j-1})$ .
2. Festlegung einer Fortsetzungsschrittweite  $h_j > 0$  und des neuen Parameterwertes  $\lambda_j = \lambda_{j-1} + h_j$ .
3. Prädiktorschritt: Vorgabe eines Näherungswertes  $x^P$  für den neuen Lösungsvektor  $x_j = x(\lambda_j)$ .



**Abb. 2** Fortsetzung mit natürlicher Parametrisierung

4. Korrektorschritt: Iterative Verbesserung des Prädiktorwertes mittels eines Newton-ähnlichen Verfahrens für Gleichung (21).

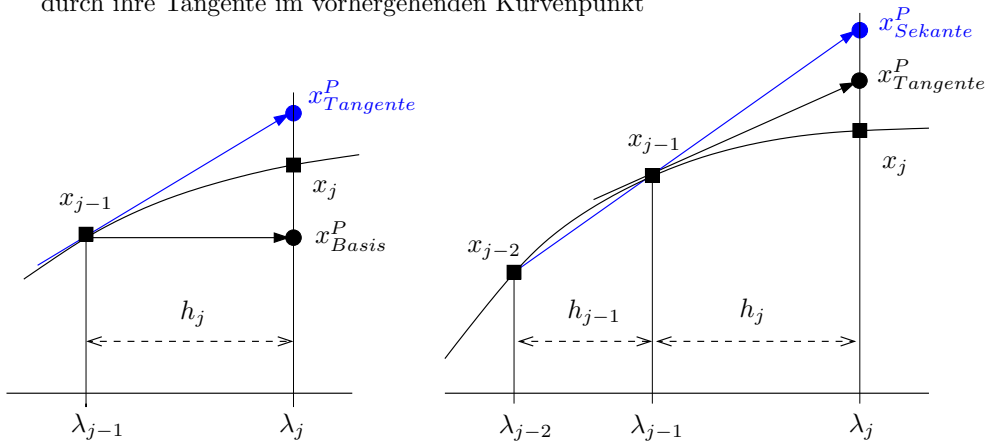
### Prädiktorschritt

Für den *Prädiktorschritt* nutzen wir Informationen über zuvor bestimmte Kurvenpunkte und extrapolieren die Lösungskurve bis zu  $\lambda_j$ :

1. Der *Basis-Prädiktor* nimmt den vorhergehenden Kurvenpunkt als Startwert für den folgenden Korrektorschritt:

$$x^P := x_{j-1} = x(\lambda_{j-1}) \quad (26)$$

2. Der *Tangenten-Prädiktor* – auch *Euler-Prädiktor*<sup>2</sup> genannt – ersetzt die Kurve lokal durch ihre Tangente im vorhergehenden Kurvenpunkt



**Abb. 3** Prädiktoren bei natürlicher Parametrisierung

<sup>2</sup>Leonhard Euler (1707–1783), Schweizer Mathematiker, arbeitete auf fast allen Gebieten der Mathematik; Leistungen zur Mechanik, Technik, Optik und Astronomie.

$$x(\lambda) = x(\lambda_{j-1}) + (\lambda - \lambda_{j-1}) \cdot x'(\lambda_{j-1}), \quad (27)$$

wobei  $x'(\lambda_{j-1})$  mit der Einbettungsgleichung (24) gewonnen werden kann. Wegen  $x_{j-1} := x(\lambda_{j-1})$  ergibt sich so die Prädiktorformel

$$x^P := x_{j-1} - h_j \cdot [D_x f(x_{j-1}, \lambda_{j-1})]^{-1} D_\lambda f(x_{j-1}, \lambda_{j-1}).$$

Sie besitzt wie in Abbildung 3 links dargestellt eine höhere Genauigkeit als (26), erfordert allerdings die Ableitungen  $D_x f$  und  $D_\lambda f$  sowie die Lösung eines linearen Gleichungssystems.

3. Legt man durch die zwei Knoten  $(x_{j-2}, \lambda_{j-2})$  und  $(x_{j-1}, \lambda_{j-1})$  die Sekante und extrapoliert bis zum Parameterwert  $\lambda_j$ , so liefert dieser *Sekanten-Prädiktor* den in Abbildung 3 rechts dargestellten Wert

$$x^P := x_{j-1} + \frac{\lambda_j - \lambda_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} \cdot [x_{j-1} - x_{j-2}]. \quad (28)$$

Er benötigt für  $x^P$  bereits an der Stelle  $\lambda_1$  zwei Startwerte, die man sich z. B. mittels des Basis-Prädiktors verschaffen muss. Dieser Ansatz stellt einen praktikablen Zugang dar.

4. Durch Nutzung weiterer zurückliegender Kurvenpunkte  $(x_{j-1}, \lambda_{j-1}), (x_{j-2}, \lambda_{j-2}), \dots, (x_{j-k}, \lambda_{j-k})$  kann man mittels *polynomialer Extrapolation* sehr genaue Prädiktorformeln der allgemeinen Form

$$x^P := \alpha_{1j} x_{j-1} + \alpha_{2j} x_{j-2} + \dots + \alpha_{kj} x_{j-k}, \quad \alpha_{ij} \in \mathbb{R}, \quad (29)$$

gewinnen und den Sekanten-Prädiktor weiter verbessern. Mit verfügbaren Ableitungen  $x'(\lambda_{j-1}), x'(\lambda_{j-2}), \dots, x'(\lambda_{j-k})$  lassen sich mit *Hermite-Interpolation* noch genauere Prädiktorwerte  $x^P$  ermitteln (vgl. Deuffhard (2004)).

Um die Genauigkeit dieser und weiterer Prädiktoren zu vergleichen, betrachten wir  $x(\lambda)$  an der Stelle  $\lambda = \lambda_{j-1}$  und definieren mit  $\lambda = \lambda_{j-1} + h$  die lokale Funktion

$$u(h) := x(\lambda) = x(\lambda_{j-1} + h) \quad (30)$$

für  $h \in I_\varepsilon = [-\varepsilon, \varepsilon], \varepsilon > 0$ . Die drei angeführten Prädiktoren lassen sich dann ebenfalls in Abhängigkeit von  $h$  angeben:

1. Basis-Prädiktor (26):  $u_B(h) = x_{j-1}$
2. Tangenten-Prädiktor (27):  $u_T(h) = x_{j-1} + h \cdot u'(0)$
3. Sekanten-Prädiktor (28):  $u_S(h) = x_{j-1} + h \cdot (x_{j-1} - x_{j-2}) / (\lambda_{j-1} - \lambda_{j-2})$ .

Wir führen nun eine Approximationsordnung  $p$  ein:  $u(h)$  und  $u_P(h)$  seien zwei gegebene Funktionen mit  $u, u_P : I_\varepsilon \rightarrow \mathbb{R}^n$ . Dann approximiert  $u_P(h)$  die Kurve  $u(h)$  mit der Ordnung  $p \in \mathbb{N}$  an der Stelle  $h = 0$ , falls es Konstanten  $0 < h_0 \leq \varepsilon$  und  $K > 0$  gibt, so dass

$$\|u(h) - u_P(h)\| \leq K \cdot |h|^p \quad \forall |h| < h_0 \quad (31)$$

gilt, d. h.  $u(h) - u_P(h) = O(|h|^p)$  ist. Für die drei Prädiktoren beweist man dann

### Satz 25

Unter Voraussetzung 23 approximieren der Basis-Prädiktor und der Sekanten-Prädiktor mit Ordnung 1, während der Tangenten-Prädiktor mit Ordnung 2 approximiert.

**Beweis:** Es sei  $|h| < h_0$ . Für die drei Prädiktoren erhalten wir

$$\begin{aligned} \text{(i)} \quad \|u(h) - u_K(h)\| &= \|u(h) - x_{j-1}\| = \|u(h) - u(0)\| \\ &= \left\| \int_0^1 u'(th) dt \cdot h \right\| \leq \int_0^1 \|u'(th)\| dt \cdot |h| \\ &\leq K_B \cdot |h| \quad \text{mit} \quad K_B := \max_{t \in [-\varepsilon, \varepsilon]} \|u'(t)\|. \\ \text{(ii)} \quad \|u(h) - u_T(h)\| &= \|u(h) - x_{j-1} - h \cdot u'(0)\| \\ &= \|u(0) + u'(0) \cdot h + \frac{1}{2} u''(\sigma) \cdot h^2 - u(0) - u'(0) \cdot h\| \\ &\leq \frac{1}{2} \max_{t \in [-\varepsilon, \varepsilon]} \|u''(t)\| \cdot |h|^2 = K_T \cdot |h|^2. \\ \text{(iii)} \quad \|u(h) - u_S(h)\| &= \|u(h) - u(0) - h \cdot (x_{j-1} - x_{j-2}) / (\lambda_{j-1} - \lambda_{j-2})\| \\ &\leq \|u(h) - u(0)\| + \|x'(\lambda^*) \cdot h\| \quad \text{mit} \quad \lambda^* \in (\lambda_{j-1}, \lambda_{j-2}) \\ &\leq K_1 \cdot |h| + K_2 \cdot |h| = K_S \cdot |h| \end{aligned}$$

mit  $K_S := K_1 + K_2$ ,  $K_1 := \max_{t \in [-\varepsilon, \varepsilon]} \|u'(t)\|$ ,  $K_2 := \max_{\lambda \in [\lambda_{j-2}, \lambda_{j-1}]} \|x'(\lambda)\|$ .  $\square$

Berücksichtigen wir auch  $(x_{j-3}, \lambda_{j-3})$  und interpolieren mit den Punkten  $(x_{j-3}, \lambda_{j-3})$ ,  $(x_{j-2}, \lambda_{j-2})$ ,  $(x_{j-1}, \lambda_{j-1})$ , so gewinnen wir einen *Parabel-Prädiktor*, der unter der Voraussetzung  $f \in C^3(D \times \Lambda)$  ebenfalls die Approximationsordnung 2 besitzt.

### Korrektorschritt

Wegen seiner schnellen Konvergenz wird für den *Korrektorschritt* häufig das Newton-Verfahren (5) genutzt. Mit dem berechneten Prädiktorpunkt  $(x^P, \lambda_j)$  setzen wir  $u_0 := x^P$  und gewinnen formal die Näherungslösungen

$$u_{k+1} := u_k - [D_x f(u_k, \lambda_j)]^{-1} f(u_k, \lambda_j), \quad k = 0, 1, 2, \dots \quad (32)$$

Numerisch effektiver ist auch hier die Lösung linearer Gleichungssysteme

$$D_x f(u_k, \lambda_j) \cdot \Delta u_k = -f(u_k, \lambda_j), \quad u_{k+1} := u_k + \Delta u_k, \quad k = 0, 1, 2, \dots, K-1. \quad (33)$$

Der ermittelte Punkt  $(x_j, \lambda_j)$  mit dem Wert  $x_j := u_K$  liegt im Allgemeinen nicht auf der Kurve  $\mathcal{L}$ , da der Abbruch des Verfahrens bereits nach endlich vielen Iterationen erfolgt. Ein geeignetes Abbruchkriterium muss deshalb vorgesehen werden. Um die aufwändigen Berechnungen der Jacobi-Matrizen zu vermeiden, werden oft Newton-ähnliche Verfahren mit konstanter Matrix  $A_0 \sim D_x f(x^P, \lambda_j)$  bevorzugt, so dass (33) nun die Form

$$A_0 \cdot \Delta u_k = -f(u_k, \lambda_j), \quad u_{k+1} := u_k + \Delta u_k \quad (34)$$

erhält, z. B. mit einer Differenzenapproximation der Matricelemente

$$\{A_0\}_{ij} := \frac{1}{\Delta x_j} \{f_i(x_1^P, \dots, x_j^P + \Delta x_j, \dots, x_n^P, \lambda) - f_i(x^P, \lambda)\}, \quad i, j = 1(1)n. \quad (35)$$

Die Verifikation des Fortsetzungsverfahrens erfordert den Nachweis der lokalen Konvergenz dieses Korrektorschrittes bei geeignetem Prädiktorwert. Damit ist die Endlichkeit der Anzahl von Fortsetzungsschritten über dem Parameterintervall  $\Lambda = [a, b]$  zu begründen. Bei Deuffhard und Hohmann (1993) findet man dazu folgendes

**Lemma 26 (Fortsetzungsschritt)**

*Zu  $\varepsilon > 0$  sei  $\Lambda_\varepsilon = (a - \varepsilon, b + \varepsilon)$  ein offenes Intervall, das  $\Lambda = [a, b]$  enthält. Voraussetzung 23 sei mit  $\Lambda_\varepsilon$  erfüllt. Weiterhin existiere ein  $\omega > 0$ , mit dem  $f$  der Lipschitz-Bedingung*

$$\|D_x f(x, \lambda)^{-1} \{D_x f(x + hv, \lambda) - D_x f(x, \lambda)\}v\| \leq h\omega \|v\|^2 \quad (36)$$

*mit  $(x, \lambda) \in D \times \Lambda_\varepsilon$ ,  $x + hv \in D$ , genügt. Das Prädiktorverfahren  $P$  besitze die Ordnung  $p$ , d. h. es existiert  $K > 0$  mit*

$$\|u(h) - u_P(h)\| \leq K \cdot |h|^p \quad \forall |h| < \varepsilon \quad (37)$$

*mit  $u(h) = x(\lambda_{j-1} + h)$ ,  $\lambda_{j-1} \in \Lambda$  und Prädiktorwert  $u_P(h)$ . Dann konvergiert das Korrekturverfahren (33) mit Startwert  $u_0 := x_j^P = u_P(h_j)$  gegen die Lösung  $u(h_j) = x(\lambda_j)$  von*

$$f(x, \lambda_j) = f(x, \lambda_{j-1} + h_j) = 0, \quad (38)$$

*falls  $0 < h_j < h_{max} := \max \left\{ \varepsilon, \sqrt[2]{2/(K\omega)} \right\}$  ist.*

Wechselweise Anwendung des Prädiktor- und Korrektorschrittes zusammen mit der Erhöhung des Parameterwertes  $\lambda_j$  ergibt den Algorithmus 28 für das *Prädiktor-Korrektor-Verfahren (PECE)* zur numerischen Lösungsfortsetzung. Um dessen unbeschränkte Ausführbarkeit zu verifizieren, ist eine Auswahlregel für die Schrittweiten  $h_j$  anzugeben, die die Durchführbarkeit und Endlichkeit der Fortsetzung sichert.



**Satz 27 (PECE-Verfahren)**

Die Voraussetzungen des Lemmas 26 seien erfüllt. Für  $\lambda_0 = a$  genüge der Startwert  $x_0 \approx x(\lambda_0)$  der Bedingung  $\rho := \|x_0 - x(\lambda_0)\| < 2/\omega$  und  $K[x(\lambda_0); \rho] \subseteq D$ . Wählt man die Schrittweiten  $h_j$  mit  $h_j := \lambda_j - \lambda_{j-1}$  gemäß

$$h_j := \min\{h_{\max}/2, b - \lambda_{j-1}\}, \quad (39)$$

so gelten folgende Behauptungen:

- i. Algorithmus 28 ist durchführbar, d. h. es existieren die  $\lambda_j$  und  $x(\lambda_j)$ ,  $j = 0, 1, 2, \dots$  mit  $x(\lambda_j)$  lokal eindeutige Lösung von  $f(x(\lambda_j), \lambda_j) = 0$ ,  $j = 0, 1, 2, \dots$
- ii. Algorithmus 28 ist endlich, d. h.  $\lambda_n = b$  für ein  $N \in \mathbb{N}$ .

Die Endlichkeit des Korrektorschrittes wird damit allerdings nicht gesichert (vgl. Schwetlick (1979)). Wir verzichten auf den Induktionsbeweis und verweisen auf die sehr guten Darstellungen bei Deuffhard und Hohmann (1993) sowie Deuffhard (2004).

**Schrittweitensteuerung**

Die richtige Wahl der Fortsetzungsschrittweite  $h_j = \lambda_j - \lambda_{j-1}$  ist oft entscheidend für den Erfolg der Fortsetzungsmethode. Bei zu großem Wert kann der Prädiktorpunkt außerhalb des Einzugsbereiches der zu bestimmenden Lösung liegen. Dies führt meist zur Divergenz des Korrektors, mitunter auch zur Konvergenz gegen eine andere unerwünschte Lösung. Die theoretisch korrekte Wahl (39) führt oft auf extrem kleine Schrittweiten  $h_j$ , womit sich die Zahl  $N$  der Fortsetzungsschritte und somit der Gesamtaufwand über alle Maßen erhöht. Eine einfache aber zuverlässige Strategie soll solche  $\lambda$ -Schrittweiten  $h = h_j > 0$  wählen, die die Konvergenz des Korrektors gegen die verfolgte Lösung garantiert und zugleich die Anzahl der Rechenoperationen minimiert. Heuristisch lässt sich der von  $h$  abhängige Rechenaufwand  $A(h)$ , bezogen auf einen Einheits-Fortsetzungsschritt, durch die Anzahl

$$A(h) := k(h)/h \quad \text{bei } k(h) \text{ ausgeführten Korrektorschritten}$$

ansetzen. Wird mit einer „optimalen“ Schrittweite  $h_{\text{opt}} = \varrho \cdot h$  der minimale Aufwand

$$A(h_{\text{opt}}) := k(h_{\text{opt}})/h_{\text{opt}}$$

erreicht, so liefert Division mit den Abkürzungen  $k := k(h)$ ,  $k_{\text{opt}} := k(h_{\text{opt}})$

$$\frac{A(h)}{A(h_{\text{opt}})} = \frac{k \cdot h_{\text{opt}}}{h \cdot k_{\text{opt}}} = \frac{k}{k_{\text{opt}}} \cdot \varrho \geq 1.$$

Also ist  $\varrho \geq k_{\text{opt}}/k$  sinnvoll. Geben wir eine optimale Anzahl von Korrektoriterationen  $k_{\text{opt}}$  vor, so setzen wir deshalb heuristisch

$$\rho := k_{\text{opt}}/k. \quad (40)$$

**ALGORITHMUS 28 (Natürliche Parametrisierung)**Function  $[N, \lambda, x] = \text{continuation}(f, Df x, a, b, x_0, \text{tol})$ 

1. Setze  $j := 0$ ,  $\lambda_0 := a$ ,  $x^P := x_0$   
Wähle Schrittweite  $h$  sowie Iterationszahl  $k_{\text{opt}}$ .
2. (Startrechnung) Bestimme  $x_0 = x(\lambda_0)$  mit Verfahren (33) oder (34).
3. Do while  $\lambda_j \leq b$ 
  1. Setze  $j := j + 1$ .
  2. Setze  $\lambda_j := \lambda_{j-1} + h$ . Falls  $\lambda_j > b$ , so  $\lambda_j := b$ .
  3. (Prädiktor) Bestimme  $x^P$  mit Verfahren (26), (27) oder (28).
  4. (Korrektor) Bestimme  $x_j = x(\lambda_j)$  mit Verfahren (33) oder (34)  
in  $k$  Schritten bis auf eine Genauigkeit  $\text{tol}$ .
  5. (Schrittweiten-Bestimmung)
    - (a) Bestimme  $\rho := k_{\text{opt}}/k$ .
    - (b) Beschränke  $\rho := \max\{\min(\rho, 2), \frac{1}{2}\}$ .
    - (c) Setze  $h := h * \rho$ .
    - (d) Falls  $h < \varepsilon_M$  (Maschinengenauigkeit), so STOP.
4. Return  $N := j$ ,  $\lambda_j$ ,  $x_j$ ,  $j = 0(1)N$

Eine zusätzliche Begrenzung  $\frac{1}{2} \leq \rho \leq 2$  (safeguarding) erhöht die Sicherheit der Wahl von  $h_{\text{opt}} = \rho \cdot h$ . Wir erhalten so die Schrittweitensteuerung in Algorithmus 28. Sie passt die aktuelle Fortsetzungsschrittweite  $h_j$  nach jedem Schritt so an, dass eine „optimale“ Zahl  $k_{\text{opt}}$  von Korrektorschritten ausgeführt wird. Für das Newton-Verfahren hat sich  $k_{\text{opt}} = 5$  oder  $k_{\text{opt}} = 6$  bewährt. Bei zu geringer Fortsetzungsschrittweite  $h$ , insbesondere bei Unterschreiten der relativen Maschinengenauigkeit  $\varepsilon_M$ , sollte der Algorithmus definiert abgebrochen werden.

Konvergiert der Korrektorschritt nicht, so versagt diese „a-posteriori“-Steuerung allerdings. Deshalb wurden zusätzliche Strategien entwickelt, die nach dem Start des Korrektors dessen Schrittweite „a-priori“ neu bestimmen und den Korrektor wiederholen können. Ein Zugang von Allgower und Georg (1990) wertet das asymptotische Verhalten von Prädiktor und Korrektor aus. Nach Satz 49 hat der Tangenten-Prädiktor  $x^P$  die Ordnung 2 bezüglich der Fortsetzungsschrittweite  $h$ . Benutzen wir als Korrektor das Newton-Verfahren, so liefert (32) den ersten Näherungswert

$$x^N := x^P - [D_x f(x^P, \lambda_j)]^{-1} f(x^P, \lambda_j).$$

Nach Ausführung eines vereinfachten Newton-Schrittes mit  $D_x f(x^P, \lambda_j)$  ergibt sich auf Grund derselben LU-Zerlegung der Jacobi-Matrix mit geringem Zusatzaufwand

$$x^V := x^N - [D_x f(x^P, \lambda_j)]^{-1} f(x^N, \lambda_j).$$

Aus beiden Werten lässt sich eine Näherung für die lineare Konvergenzrate

$$\kappa := \frac{\|x^V - x^N\|}{\|x^N - x^P\|} = \frac{\|D_x f(x^P, \lambda_j)^{-1} f(x^N, \lambda)\|}{\|D_x f(x^P, \lambda_j)^{-1} f(x^P, \lambda)\|}$$

ermitteln. Gesucht ist die explizite Abhängigkeit des  $\kappa$  von der gewählten Schrittweite  $h > 0$ , für  $h \rightarrow 0$ . Allgower und Georg (1990) weisen nach, dass unter den obigen Voraussetzungen eine von  $h$  unabhängige Konstante  $K(x^P) > 0$  existiert, so dass

$$\kappa = K(x^P) \cdot h^2 + O(h^3), \quad h < h_0 \quad (41)$$

gilt. Zu gegebenem Prädiktorwert  $x^P$  steht uns die Kontraktionsrate  $\kappa_{\text{alt}} = \kappa$  mit dieser Asymptotik zur Verfügung. Gefordert wird eine neue Kontraktionsrate  $\kappa_{\text{neu}}$  mit

$$\kappa_{\text{neu}} = K(x^P) \cdot h_{\text{neu}}^2 + O(h_{\text{neu}}^3), \quad h_{\text{neu}} < h_0. \quad (42)$$

Der Schrittweiten-Faktor sei wieder  $\rho := h_{\text{neu}}/h$ . Division der Kontraktionsraten liefert wegen  $h_{\text{neu}} = \rho \cdot h$

$$\frac{\kappa_{\text{neu}}}{\kappa_{\text{alt}}} = \frac{K(x^P) \cdot h_{\text{neu}}^2 + O(h_{\text{neu}}^3)}{K(x^P) \cdot h^2 + O(h^3)} = \left( \frac{h_{\text{neu}}}{h} \right)^2 + O(h),$$

woraus sich nach Vernachlässigung der Terme  $O(h)$  die einfache Regel für

$$h_{\text{neu}} := h \cdot \sqrt{\frac{\kappa_{\text{neu}}}{\kappa_{\text{alt}}}} \quad \text{bzw.} \quad \rho := \sqrt{\frac{\kappa_{\text{neu}}}{\kappa_{\text{alt}}}}. \quad (43)$$

ergibt. In praxi setzt man häufig  $\kappa_{\text{neu}} := 1/2$ . Bei Benutzung des Sekantenprädiktors lässt sich unter geeigneten Voraussetzungen ebenfalls eine asymptotische Darstellung

$$\kappa = K(x^P) \cdot h + O(h^2) \quad \text{mit Konstante} \quad K(x^P) > 0 \quad (44)$$

nachweisen; folglich ist

$$h_{\text{neu}} := h \cdot \frac{\kappa_{\text{neu}}}{\kappa_{\text{alt}}}. \quad (45)$$

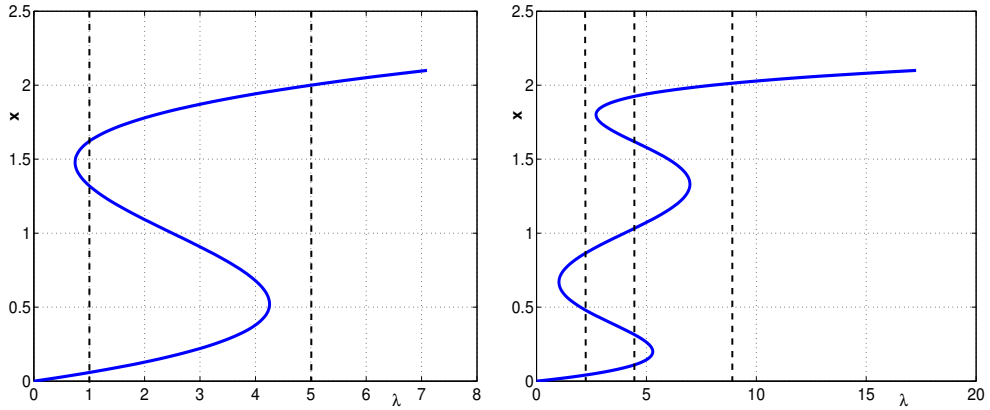
Eine Implementation dieser verbesserten Schrittweiten-Strategie wird empfohlen, da sie nicht unwesentlich zur numerischen Stabilität von Algorithmus 28 beiträgt.

### Beispiel 29

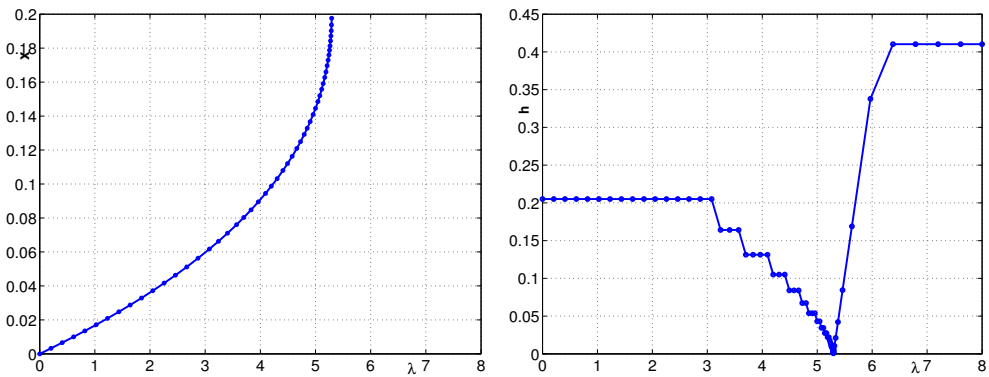
Betrachten wir das eindimensionale Beispiel  $f(x, \lambda) = 0$  aus Abbildung 4 mit

$$f(x, \lambda) := 40\mu x^5 - 200\mu x^4 + (342\mu + 8)x^3 - (226\mu + 24)x^2 + 0.5(91\mu + 37)x - \lambda,$$

wobei  $\lambda \geq 0$  der Systemparameter ist. Für festen Parameter  $\mu = 0$  treten z. B. bei  $\lambda = 1$  drei Lösungen (links), für  $\mu = 1$  sogar fünf Lösungen (rechts) auf. Wegen  $D_\lambda f(x, \lambda) = -1$  ist jede Lösung  $y = (x, \lambda)$  von  $f(y) = 0$  ein regulärer Punkt von  $f$  im Sinne der Definition 21. Die natürliche Parametrisierung nach Algorithmus 28 liefert mit Startwerten  $(x_0, \lambda_0) = (0, 0)$  die Lösung zuverlässig bis zum ersten Umkehrpunkt (vgl. Abbildung 5 links). Dann bricht das Verfahren mit extremer Schrittweitenreduktion zusammen – wie in Abbildung 5 rechts dargestellt. ■



**Abb. 4** Exakte Lösungskurve  $\mathcal{L}$  zu Beispiel 29 mit  $\mu = 0$  (links) und  $\mu = 1$  (rechts)



**Abb. 5** Berechnete Lösung mit  $\mu = 1$  (links) und Arbeits-Schrittweite  $h$  (rechts)

### 3.2 Bogenlängen-Parametrisierung

Obwohl in jedem Punkt  $y = (x, \lambda)$  des Beispiels 29 die vollständige Jacobi-Matrix  $f'(y) = (D_x f(y), D_\lambda f(y))$  maximalen Rang  $n$  hat, versagt der Fortsetzungsalgorithmus an den Stellen  $y_0 = (x_0, \lambda_0)$ , an denen  $D_x f(x_0, \lambda_0)$  singularär wird. Betrachten wir das unterbestimmte System (22) und treffen nun folgende

**Voraussetzung 30**

$D \subset \mathbb{R}^{n+1}$  sei offen.

- i.  $f \in C^r(D)$ ,  $r \geq 2$
- ii. Es existiert ein  $y_0 \in D \subset \mathbb{R}^{n+1}$  mit  $f(y_0) = 0$ .
- iii. Die erweiterte Jacobi-Matrix  $f'(y_0) = (D_x f(y_0), D_\lambda f(y_0))$  hat maximalen Rang  $n$ .

Bedingung iii. schwächt die bisherige Regularitätsvoraussetzung ab. Sie garantiert jedoch (vgl. dazu Satz 2.47 bei Marx und Vogt (2010)), dass die Menge aller Lösungen der Gleichung (22) eine eindimensionale Untermannigfaltigkeit  $\mathcal{L}$  des  $\mathbb{R}^n$  bildet. Aus dem impliziten Funktionentheorem (vgl. Hoffmann et al. (2005)) gewinnen wir dann

**Satz 31**

Unter Voraussetzung 30 existiert eine glatte Kurve  $\mathcal{L} \in \mathbb{R}^{n+1}$  mit Parametrisierung  $y = y(s)$ ,  $s \in J = (s_0 - \delta, s_0 + \delta)$ ,  $\delta > 0$ ,  $y : J \longrightarrow \mathbb{R}^{n+1}$ , so dass für alle  $s \in J$

i.  $y(s_0) = y_0$ , ii.  $f(y(s)) = 0$ , iii.  $\text{rang}(f'(y(s))) = n$ , iv.  $y'(s) \neq 0$  gilt.

$\mathcal{L}$  heißt *Lösungspfad* mit der Parametrisierung  $y = y(s)$ . Einsetzen der Parametrisierung in die Gleichung und Differenziation nach  $s$  liefert  $f'(y(s)) \cdot y'(s) = 0$ . Also liegt der Vektor  $y'(s)$  im eindimensionalen Nullraum  $N(f'(y(s)))$  und ist orthogonal zu allen  $n$  Zeilenvektoren dieser Matrix. Wegen der Rangaussage iii. des Satzes ist folglich die um den Zeilenvektor  $y'(s)^T$  erweiterte Jacobi-Matrix

$$H'(y(s)) := \begin{pmatrix} f'(y(s)) \\ y'(s)^T \end{pmatrix} \quad \text{mit} \quad N(f'(y(s))) = \text{span}\{y'(s)\} \quad (46)$$

regulär auf dem Pfad  $\mathcal{L}$  für  $s \in J$ . Ein Lösungspfad  $\mathcal{L} = \{y(s) \mid f(y(s)) = 0 \ \forall s \in J\}$  mit der Eigenschaft  $\text{rang}(f'(y(s))) = n$  für alle  $s \in J$  wird oft als *regulär* bezeichnet. Zur Klassifikation aller Kurvenpunkte  $y \in \mathcal{L}$  beweist man leicht

**Satz 32**

Die Bedingung  $\text{rang}(f'(y)) = n$  ist genau dann erfüllt, wenn

- i.  $D_x f(x, \lambda)$  regulär ist oder
- ii.  $\dim N(D_x f(x, \lambda)) = 1$  und  $D_\lambda f(x, \lambda) \notin R(D_x f(x, \lambda))$  gilt.

Ein Kurvenpunkt  $y(s_0) = (x(s_0), \lambda(s_0)) \in \mathcal{L}$  wird im Falle ii. als *Umkehrpunkt* (*einfacher Grenzpunkt*, *turning point*, *simple limit point*) bezeichnet. Unter Voraussetzung 30 besteht jeder reguläre Lösungspfad deshalb nur aus Punkten mit regulärer Jacobi-Matrix  $D_x f(x(s_0), \lambda(s_0))$  und einfachen Grenzpunkten. Echte Singularitäten der Ordnung  $k \geq 1$  treten hier nicht auf.

Welche praktikablen Parametrisierungen  $y = y(s)$  der Kurve  $\mathcal{L}$  sind nun einsetzbar? Eine auf Chua und Ushida (1976) zurückgehende Idee führt einen *Parameterwechsel* (*parameter switching*) in der Umgebung von Umkehrpunkten durch. Der natürliche Parameter  $\lambda = y_{n+1}$  wird dabei lokal durch eine „geeigneter“ andere Komponente  $y_p$ ,  $p \neq n+1$ , des Vektors  $y$  ersetzt. Offenbar ist  $y_p$  besonders dann geeignet, wenn sich diese Komponente beim Durchlaufen des Lösungspfades lokal am stärksten ändert. Benutzt man den Tangenten-Prädiktor  $y'(s_j)$  mit  $\|y'(s_j)\|_s = 1$ , kann man den Index  $p$  so bestimmen, dass

$$|y'_p(s_j)| = \max_{i=1(1)n+1} |y'_i(s_j)|$$

ist, während beim Sekanten-Prädiktor zu gegebenen Punkten  $y(s_{j-1})$  und  $y(s_j)$

$$|y_p(s_j) - y_p(s_{j-1})| = \max_{i=1(1)n+1} |y_i(s_j) - y_i(s_{j-1})|$$

der Auswahl von  $p$  dient. Dieser Parameterwechsel kann nach jedem Fortsetzungsschritt erfolgen, womit die Sonderrolle des Parameters  $\lambda$  nach dem ersten Wechsel mit  $y_p$  vollkommen entfällt. Wählt man zu  $p$  den neuen Parameter  $\eta := y_p$ , so erhält man das bestimmende System mit  $n$  Gleichungen in den  $n$  Variablen  $y_1, \dots, y_{p-1}, y_{p+1}, \dots, y_{n+1}$

$$f(y_1, \dots, y_{p-1}, \eta, y_{p+1}, \dots, y_{n+1}) = 0 \quad (47)$$

zu gegebenem Ausgangspunkt  $y^j = (y_1^j, \dots, y_{p-1}^j, \eta^j, y_{p+1}^j, \dots, y_{n+1}^j)$  und  $\eta^j := y_p^j$ . Nachteilig an dieser Vorgehensweise erweist sich die Gleichbehandlung aller  $n+1$  Komponenten von  $y$  bei der Wahl von Index  $p$ . Wegen unterschiedlicher Größenordnungen der  $x_i$ -Komponenten und des Parameters  $\lambda$  sollte deshalb die absolut-relative Abweichung

$$d_p := \max_{i=1(1)n+1} d_i \quad \text{mit} \quad d_i := \frac{|y_i(s_j) - y_i(s_{j-1})|}{|y_i(s_j)| + 1}, \quad i = 1(1)n+1 \quad (48)$$

betrachtet werden, die so als gewichtetes Maximum interpretierbar ist. Wir wollen die theoretische Rechtfertigung dieser Strategie jedoch nicht weiter verfolgen, sondern die auf Doedel et al. (2002) zurückgehende Idee der „Pseudo-Bogenlänge“ präferieren, die diesen abrupten Parameterwechsel vermeidet und sich vielfach bewährt hat.

Führen wir – ggf. durch Umparametrisierung der Kurve  $\mathcal{L}$  – die Bogenlänge  $s$  von  $[0, \sigma]$

$$s(\sigma) := \int_0^\sigma \sqrt{\sum_{i=1}^{n+1} [y'_i(\sigma)]^2} d\sigma = \int_0^\sigma \|y'(\sigma)\|_2 d\sigma$$

als Kurvenparameter ein, so genügt der Tangentenvektor in jedem Punkt der Beziehung

$$\|y'(s)\|_2 = 1 \quad \forall s \in J. \quad (49)$$

Die Durchlaufgeschwindigkeit einer Kurve ist tatsächlich genau dann in jedem Punkt  $s \in J$  gleich Eins, wenn der Parameter  $s$  gleich der Bogenlänge ist. Der Tangentenvektor  $y'(s)$  wird durch das um (49) erweiterte System

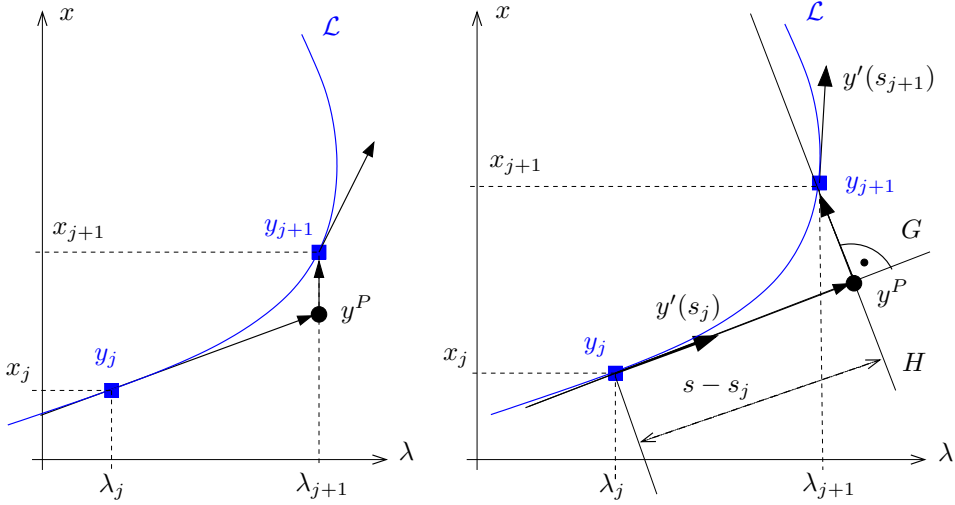
$$\left. \begin{array}{l} f'(y(s)) \cdot y'(s) = 0 \\ y'(s)^T y'(s) - 1 = 0 \end{array} \right\} \quad \text{mit} \quad \det \begin{pmatrix} f'(y(s)) \\ y'(s)^T \end{pmatrix} > 0 \quad (50)$$

eindeutig bestimmt. Allerdings sollte die numerische Auswertung dieses nichtlinearen impliziten DGL-Systems für  $y(s)$  vermieden werden. Nehmen wir dazu an, dass für  $s_{j-1}$  die Lösung  $y_{j-1} = y(s_{j-1})$  bekannt ist, d. h.  $y'(s_{j-1})^T y'(s_{j-1}) - 1 = 0$ . Linearisierung von (50) bzgl.  $y'(s)$  bei  $y_{j-1}$  durch lokale Approximation der Bogenlänge liefert dann

$$y'(s_{j-1})^T \frac{y(s) - y(s_{j-1})}{s - s_{j-1}} - 1 = 0,$$

woraus wir nach Multiplikation mit  $s - s_{j-1}$  die skalare Zusatzbedingung für  $y$

$$g(y) := y'(s_{j-1})^T (y - y_{j-1}) - (s - s_{j-1}) = 0, \quad g: \mathbb{R}^{n+1} \rightarrow \mathbb{R} \quad (51)$$



**Abb. 6** Natürliche Parametrisierung (links) und Pseudo-Bogenlänge (rechts)

erhalten. Ergebnis ist das *erweiterte System* für  $y$

$$h(y) := \begin{pmatrix} f(y) \\ g(y) \end{pmatrix} = 0, \quad h : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1} \quad (52)$$

mit der eingeführten skalarwertigen Funktion  $g(y)$ . Die Jacobi-Matrix dieses Systems

$$h'(y(s)) = \begin{pmatrix} f'(y(s)) \\ g'(y(s)) \end{pmatrix} = \begin{pmatrix} f'(y(s)) \\ y'(s_{j-1})^T \end{pmatrix} \quad (53)$$

ist regulär für  $s \in J_0 = (s_{j-1} - \delta_0, s_{j-1} + \delta_0)$ ,  $\delta_0 > 0$ . Denn nach (46) ist  $h'(y(s_{j-1}))$  regulär. Wegen der Stetigkeit von  $f'(y)$  ist dann auch  $h'(y(s))$  regulär.

Geometrisch lässt sich die Zusatzbedingung mit Abbildung 6 (rechts) leicht interpretieren: Wegen der Gleichung  $\tilde{y}(s) = y_{j-1} + (s - s_{j-1}) \cdot y'(s_{j-1})$  der Kurventangente  $G$  durch  $y_{j-1}$  hat der Prädiktorpunkt

$$y^P = y_{j-1} + (s - s_{j-1}) \cdot y'(s_{j-1}) \quad (54)$$

den Abstand  $s - s_{j-1}$  von  $y_{j-1}$ , da  $\|y'(s_{j-1})\|_2 = 1$  ist. Die Normal-Hyperebene  $H$  zu  $G$  durch  $y^P$  lautet dann

$$y'(s_{j-1})^T \cdot (y - y^P) = 0.$$

Einsetzen des  $y^P$  in  $H$  ergibt

$$0 = y'(s_{j-1})^T \{y - y_{j-1} - (s - s_{j-1})y'(s_{j-1})\} = y'(s_{j-1})^T (y - y_{j-1}) - (s - s_{j-1}) = g(y),$$

d. h.  $g(y)$  beschreibt die Normal-Hyperebene  $H$  zu  $y'(s_{j-1})$ , die den Abstand  $s - s_{j-1}$  von  $y_{j-1}$  hat. Wir fassen die Betrachtungen zusammen in

**Satz 33**

Mit den Voraussetzungen 30 sei  $y_{j-1} = y(s_{j-1}) \in \mathcal{L}$  gegeben. Dann existiert ein  $\delta > 0$ , so dass für alle  $s$  mit  $|s - s_{j-1}| < \delta$  das erweiterte System

$$\begin{aligned} f(y) &= 0 \\ g(y) = y'(s_{j-1})^T \cdot (y - y_{j-1}) - (s - s_{j-1}) &= 0 \end{aligned} \quad (55)$$

einen eindeutigen glatten Lösungszweig  $y = y(s)$  besitzt. Die Jacobi-Matrix  $h'(y(s))$  längs des Zweiges ist regulär.

**Beweis:**  $h \in C^r(D \times \Lambda)$ ,  $h(y(s_{j-1})) = 0$  und  $h'(y(s_{j-1}))$  ist invertierbar. Mit (53) folgt die eindeutige Existenz von  $y(s)$  nach dem impliziten Funktionentheorem.  $\square$

**ALGORITHMUS 34 (Pseudo-Bogenlänge)**

Function  $[N, y] = \text{pseudoarclength}(f, Df, a, b, x, \text{smax}, \text{tol})$

1. Setze  $j := 0$ ,  $\lambda_0 := a$ ,  $x^P := x$ ,  $s_0 := 0$   
Wähle Schrittweite  $h$  sowie Iterationszahl  $k_{\text{opt}}$ .
2. (*Startrechnung*) Bestimme  $x_0 = x(\lambda_0)$  mit Verfahren (33) oder (34).  
Setze  $y_0 = (x_0, \lambda_0)$ .
3. Do while ( $a \leq \lambda_j \leq b$  and  $s_j < \text{smax}$ )
  - 3.1. (*Tangentenvektor*) Bestimme  $A := f'(y_j)$  und  $y'(s_j)$ .
  - 3.2. Setze  $j := j + 1$  und  $s_j := s_{j-1} + h$ .
  - 3.3. (*Tangenten-Prädiktor*) Bestimme  $y^P$  mittels (54).
  - 3.4. (*Newton-Korrektor*) Löse das erweiterte System (55) in  $\mathbb{R}^{n+1}$   
mit dem Newton-Verfahren oder Newton-ähnlichem Verfahren
  - 3.5. (*Schrittweiten-Anpassung*) wie in Algorithmus 28
4. Return  $N := j$ ,  $y_j = (x_j, \lambda_j)$ ,  $j = 0(1)N$

Der Bogenlängen-Parameter wird in Abbildung 6 rechts anschaulich durch den Tangentenparameter  $s$  ersetzt, weshalb man nun von der *Pseudo-Bogenlänge* spricht. Anders als bei der natürlichen Parametrisierung in Abbildung 6 links werden nun Umkehrpunkte leicht überwunden. Algorithmus 34 nutzt dafür den Tangenten-Prädiktor.

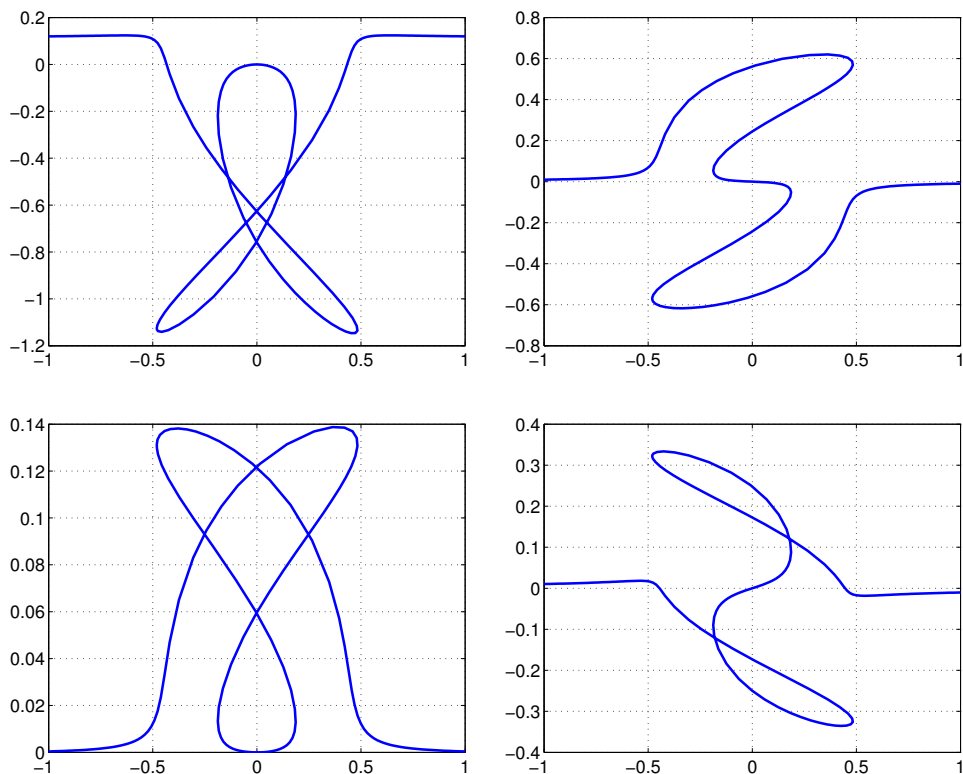


**Beispiel 35**

Deuffhard (2004) behandelt ein fünfdimensionale Stabilitätsproblem aus dem Flugzeugbau mit den Gleichungen

$$\begin{aligned}
 & -3.933x_1 + 0.107x_2 + 0.126x_3 - 9.99x_5 - 45.83\lambda - 0.727x_2x_3 \\
 & \quad + 8.39x_3x_4 - 684.4x_4x_5 + 63.5x_4\lambda = 0 \\
 & -0.987x_2 - 22.95x_4 - 28.37u + 0.949x_1x_3 + 0.173x_1x_5 = 0 \\
 & 0.002x_1 - 0.235x_3 + 5.67x_5 - 0.921\lambda - 0.713x_1x_2 - 1.578x_1x_4 + 1.132x_4\lambda = 0 \\
 & \quad x_2 - x_4 - 0.168u - x_1x_5 = 0 \\
 & -x_3 - 0.196x_5 - 0.0071\lambda + x_1x_4 = 0.
 \end{aligned}$$

Die Steuervariable  $u$  wird hier auf Null gesetzt und  $-1 \leq \lambda \leq 1$  angenommen. Algorithmus 34 liefert zuverlässig die Darstellungen der Lösungskomponenten



**Abb. 7** Darstellung der Lösungskomponenten  $x_2$  (links oben),  $x_3$  (rechts oben) und  $x_4$  (links unten),  $x_5$  (rechts unten) über dem Parameter  $\lambda$

$x_2, x_3, x_4, x_5$  in Abbildung 7 mit einer vorgegebenen Genauigkeit  $tol = 10^{-8}$ . Die Lösungssymmetrie bezüglich  $\lambda = 0$  ist gut erkennbar. ■

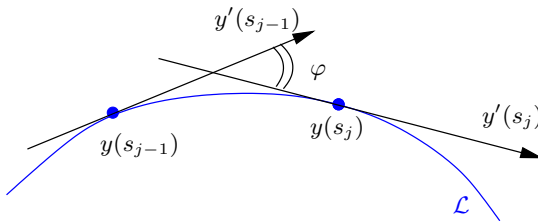
Es bleibt die Frage zu klären, wie wir den in Algorithmus 34 – sowohl im Prädiktors- als auch im Korrektorschritt – benötigten normierten Tangentenvektor  $y'(s_j)$  effektiv bestimmen. Es sei  $A := f'(y(s_j)) \in \mathbb{R}^{n \times (n+1)}$  die im  $j$ -ten Schritt ermittelte Jacobi-Matrix. Nach Voraussetzung 30 ist ihr Nullraum  $N(A)$  eindimensional. Deshalb kann  $y'(s_j) \in N(A)$  an Stelle der üblicherweise benutzten Singulärwertzerlegung in diesem speziellen Fall mittels QR-Zerlegung oder LU-Zerlegung gefunden werden. Die QR-Zerlegung der transponierten Jacobi-Matrix  $A^T$  führt auf die Darstellung

$$A^T = Q \cdot R = \begin{pmatrix} q_{11} & \cdots & q_{1,n+1} \\ q_{21} & \cdots & q_{2,n+1} \\ & \ddots & \\ q_{n+1,1} & \cdots & q_{n+1,n+1} \end{pmatrix} \cdot \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ 0 & \ddots & \\ & & r_{nn} \\ 0 & \cdots & 0 \end{pmatrix}$$

mit der orthogonalen Matrix  $Q = (q_{ik}) \in \mathbb{R}^{(n+1) \times (n+1)}$  und der hier dargestellten oberen Dreiecksmatrix  $R = (r_{ik}) \in \mathbb{R}^{(n+1) \times n}$ , deren letzte Zeile nur Nullen enthält. Mit dem  $(n+1)$ -ten Einheitsvektor  $e_{n+1} = (0, \dots, 0, 1)^T$  und der Darstellung  $z := Qe_{n+1}$  gewinnen wir die letzte Spalte der Matrix  $Q$ . Wegen der Orthogonalitätsbedingung  $Q^T Q = I$  ist

$$A \cdot z = (QR)^T Qe_{n+1} = R^T Q^T Qe_{n+1} = 0 \quad \text{und} \\ \|z\|_2^2 = e_{n+1}^T Q^T Qe_{n+1} = 1,$$

womit der Vektor  $y'(s_j) := z$  bis auf das Vorzeichen bereits das gesuchte Nullraumelement ist. Das Vorzeichen des Tangentenvektors muss nun so gewählt werden, dass ein



**Abb. 8**

Darstellung zweier aufeinander folgender Tangentenvektoren  $y'(s_{j-1})$  und  $y'(s_j)$  mit der Lösungskurve  $\mathcal{L}$

Durchlaufwechsel der Lösungskurve  $\mathcal{L}$  vermieden wird. Dazu betrachten wir zwei aufeinander folgende Tangentenvektoren  $y'(s_{j-1})$  und  $y'(s_j)$ , wie in Abbildung 8 dargestellt. Um einen spitzen Winkel  $\varphi$  zwischen beiden Tangentenrichtungen zu garantieren, fordern wir für das Skalarprodukt der Vektoren

$$\tau := y'(s_{j-1})^T y'(s_j) = \|y'(s_{j-1})\| \cdot \|y'(s_j)\| \cdot \underbrace{\cos \varphi}_{>0} > 0.$$

Falls  $\tau < 0$  ist, so kehren wir das Vorzeichen von  $y'(s_j)$  um und erhalten insgesamt den fehlenden Schritt 3.1. von Algorithmus 34:

### ■ Berechnung des Tangentenvektors

- (a) Bestimme die Jacobi-Matrix  $A := f'(y_j)$ .
- (b) QR-Zerlegung von  $A^T$  liefert  $A^T = Q \cdot R$
- (c) Setze  $y'(s_j) := Qe_{n+1}$  mit  $(n+1)$ -tem Einheitsvektor  $e_{n+1}$
- (d) Falls  $\tau := y'(s_{j-1})^T y'(s_j) < 0$ , so setze  $y'(s_j) := -y'(s_j)$ .

Man überlegt sich leicht, dass in analoger Weise auch die LU-Zerlegung von  $A^T$  mit Zeilenvertauschungen zur Ermittlung des Tangentenvektors  $y'(s_j)$  genutzt werden kann.

## 3.3 Gauß-Newton-Fortsetzung

Betrachten wir nochmals Abbildung 6, so erkennen wir, dass bei zu groß gewählter Fortsetzungsschrittweite  $h$  auch die Pseudo-Bogenlänge keinen Schnittpunkt  $y_j$  der Normal-Hyperbene mit der Kurve liefern muss. Daher sind folgende Schritte naheliegend, um vom Bogenlängen-Parameterwert  $s_{j-1}$  zu  $s_j$  überzugehen:

1. Gegeben sei der Kurvenpunkt  $y_{j-1} = y(s_{j-1})$ . Festlegung einer geeigneten Fortsetzungsschrittweite  $h_j > 0$  und des neuen Parameterwertes  $s_j = s_{j-1} + h_j$ .
2. Vorgabe eines Prädiktorpunktes  $v_0 = v = y^P$  für den neuen Kurvenpunkt  $y_j$ , z. B. mit dem Tangenten-Prädiktor

$$y^P := y_{j-1} + h_j \cdot y'(s_{j-1}), \quad \|y'(s_{j-1})\|_2 = 1 \quad (56)$$

oder dem Sekanten-Prädiktor (vgl. Abb. 9)

$$y^P := y_{j-1} + \frac{h_j}{h_{j-1}} \cdot (y_{j-1} - y_{j-2}). \quad (57)$$

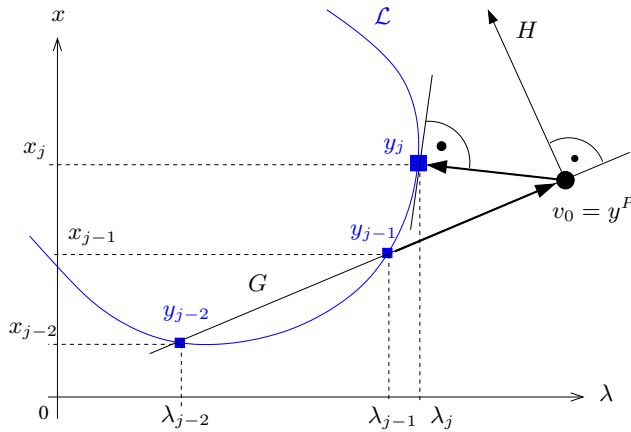
3. Bestimmung eines Kurvenpunktes  $y_j \in \mathcal{L}$  mit minimalem Abstand von  $v = y^P$ , d. h. iterative Lösung der Minimierungsaufgabe im Korrektorschritt (vgl. Abb. 9)

$$\|y_j - v\|_2 = \min_{f(y)=0} \|y - v\|_2. \quad (58)$$

Um dieses Extremwertproblem mit Nebenbedingung zu lösen, definieren wir mit den Multiplikatoren  $\mu = (\mu_1, \dots, \mu_n)^\top$  die zugehörige Lagrange-Funktion

$$L(y, \mu) = \frac{1}{2} \|y - v\|_2^2 - \mu^\top f(y) = \frac{1}{2} \sum_{i=1}^{n+1} (y_i - v_i)^2 - \sum_{i=1}^n \mu_i f_i(y). \quad (59)$$

Durch Differentiation gewinnen wir die notwendige Bedingung an  $y_j$  für ein lokales Minimum in Komponentenform



**Abb. 9** Kurvenverfolgung mittels Projektion

$$\begin{aligned}
 (i) \quad \frac{\partial L}{\partial \mu_k} &= f_k(y) = 0, \quad k = 1(1)n, \\
 (ii) \quad \frac{\partial L}{\partial y_l} &= (y_l - v_l) - \sum_{i=1}^n \mu_i \frac{\partial f_i}{\partial y_l} = 0, \quad l = 1(1)n + 1
 \end{aligned}$$

bzw. in Vektorschreibweise

$$\begin{aligned}
 (i) \quad f(y) &= 0 \\
 (ii) \quad y - v &= [f'(y)]^\top \cdot \mu, \quad \mu \in \mathbb{R}^n.
 \end{aligned} \tag{60}$$

Multiplikation von Bedingung (ii) an der Lösung  $y_j$  mit dem Tangentenvektor  $y'(s_j)$  ergibt

$$y'(s_j)^\top \cdot (y_j - v) = [f'(y(s_j)) \cdot y'(s_j)]^\top \mu = 0$$

wegen Voraussetzung 30, womit das bestimmende System für  $y_j \in \mathbb{R}^{n+1}$  nunmehr

$$\begin{aligned}
 (i) \quad f(y) &= 0 \\
 (ii) \quad y'(s_j)^T \cdot (y - v) &= 0
 \end{aligned} \tag{61}$$

lautet. Anschaulich wird also ein Minimalpunkt  $y_j$  durch die in Abbildung 9 dargestellte Orthogonalitätsbedingung  $y'(s_j)^T \cdot (y_j - v) = 0$  mit dem Tangentialvektor  $y'(s_j)$  definiert. Linearisierung des Systems (61) am Näherungspunkt  $v \in \mathbb{R}^{n+1}$  mit zugehörigem Tangentenvektor  $v'(s_j)$  liefert das lineare Gleichungssystem für den Wert  $v_{\text{neu}}$ :

$$\begin{aligned}
 f'(v) \cdot (v_{\text{neu}} - v) &= -f(v) \\
 v'(s_j)^T \cdot (v_{\text{neu}} - v) &= 0.
 \end{aligned} \tag{62}$$

Bezeichnen wir  $A := f'(v)$ ,  $b := -f(v)$  und  $x := v_{\text{neu}} - v$  sowie den Tangentenvektor  $t(A) := v'(s_j)$ , so lautet das System (62) mit diesen Abkürzungen

$$A \cdot x = b, \quad t(A) \cdot x = 0. \quad (63)$$

$Ax = b$  beschreibt eine Gerade und die Lösung  $x^*$  von (63) ist ein Punkt mit  $t(A) \perp x$ . Geometrisch bedeutet dies, eine Lösung des unterbestimmten linearen Gleichungssystems  $Ax = b$  zu finden, die die Minimierungsaufgabe

$$\|x^*\|_2 = \min_{Ax=b} \|x\|_2 \quad (64)$$

löst. Deren Lösungen  $x^*$  kann man mit der *Pseudoinversen*  $A^+$  von  $A$  formal in der Form  $x^* = A^+b$  darstellen. Dabei wird  $A^+$  folgendermaßen erklärt:

**Definition 36 (Pseudoinverse, Moore-Penrose-Inverse)**

Es sei  $A \in \mathbb{R}^{m \times n}$ . Die *Pseudoinverse* (*Moore-Penrose-Inverse*)  $A^+ \in \mathbb{R}^{n \times m}$  von  $A$  ist durch die vier Penrose-Bedingungen eindeutig definiert:

- (i)  $AA^+A = A$
- (ii)  $A^+AA^+ = A^+$
- (iii)  $(AA^+)^\top = AA^+$
- (iv)  $(A^+A)^\top = A^+A$ .

Die geometrische Beobachtung aus Abbildung 9 wird durch folgenden Äquivalenzsatz auch theoretisch abgesichert (vgl. Allgower und Georg (1990)):

**Satz 37**

Es sei  $A \in \mathbb{R}^{n \times (n+1)}$ ,  $\text{rang}(A) = n$ , und  $\varphi = t(A)$  der normierte Tangentenvektor, bestimmt durch

$$(i) \quad A \cdot \varphi = 0, \quad (ii) \quad \|\varphi\|_2 = 1, \quad (iii) \quad \det \begin{pmatrix} A \\ \varphi^\top \end{pmatrix} > 0. \quad (65)$$

Dann sind für  $b \in \mathbb{R}^n$  und  $x \in \mathbb{R}^{n+1}$  folgende Aussagen äquivalent:

- (A1)  $Ax = b$  und  $\varphi^\top x = 0$ .
- (A2)  $x = A^+b$
- (A3)  $x$  löst die Minimierungsaufgabe (64).

Die Berechnung von  $A^+$  im allgemeinen Fall einer  $(m, n)$ -Matrix ist allerdings aufwändig und bei Fortsetzungsverfahren nicht praktikabel. Wegen  $A \in \mathbb{R}^{n \times (n+1)}$  mit den Voraussetzungen 30 des zeilenregulären Falles kann  $A^+$  jedoch als Rechtsinverse

$$A^+ = A^T(AA^T)^{-1} \quad \text{wegen} \quad AA^+ = AA^T(AA^T)^{-1} = I \quad (66)$$

direkt ausgerechnet werden. Anwendung der Pseudoinversen von  $A := f'(v)$  auf (62) ergibt somit den Newtonschritt

$$v_{\text{neu}} = v - A^+ f(v) \quad \text{mit} \quad A^+ = A^T (AA^T)^{-1},$$

dessen iterative Wiederholung das *Gauß-Newton-Verfahren für unterbestimmte nichtlineare Gleichungssysteme* liefert:

$$v_{k+1} = v_k - A^T (AA^T)^{-1} f(v_k) \quad \text{mit} \quad A = f'(v_k), \quad k = 0, 1, 2, \dots \quad (67)$$

Liegt der Prädiktorwert des Startpunktes  $v_0 = y^P$  hinreichend nahe am Lösungspfad  $\mathcal{L}$ , so kann die lokale quadratische Konvergenz dieses Verfahrens nachgewiesen werden (vgl. dazu Allgower und Georg (1990)).

### Satz 38

Für  $f : D \subset \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$  seien die Voraussetzungen 30 erfüllt. Dann existiert eine offene Umgebung  $U(\mathcal{L}) \subset \mathcal{D}$  des Lösungspfades, mit der folgende Behauptungen gelten:

i. Zu jedem  $v \in U(\mathcal{L})$  existiert genau ein  $y^* \in \mathcal{L}$  mit

$$\|y^* - v\|_2 = \min_{f(y)=0} \|y - v\|_2.$$

Die dadurch definiert Abbildung  $S : U(\mathcal{L}) \rightarrow \mathcal{L}$  ist glatt.

ii. Ist  $v_0 \in U(\mathcal{L})$ , so auch  $v_1 \in U(\mathcal{L})$ .

iii. Zu jedem  $v_0 \in U(\mathcal{L})$  konvergiert die Folge  $(v_k)$ ,  $k = 0, 1, 2, \dots$  der Iterierten gemäß (67) gegen ein  $v_\infty \in \mathcal{L}$ .

iv. Für  $v_0 \in U(\mathcal{L})$  gelten folgende Abschätzungen gleichmäßig für  $v_k \in U(\mathcal{L})$ :

$$\begin{aligned} \|v_2 - v_1\| &\leq C_1 \|v_1 - v_0\|^2 \\ \|v_\infty - v_1\| &\leq C_2 \|v_\infty - v_0\|^2 \\ \|v_1 - S(v_0)\| &\leq C_3 \|v_0 - S(v_0)\|^2 \\ \|v_\infty - S(v_0)\| &\leq C_4 \|v_0 - S(v_0)\|^2 \end{aligned}$$

Zur praktischen Umsetzung dieses Verfahrens ist die vollständige Jacobi-Matrix  $A = f'(y)$  vorzugeben. Zusammen mit dem Sekanten-Prädiktor liefert der Gauß-Newton-Korrektor den Algorithmus 39. Im Unterschied zur Parametrisierung mittels Pseudo-Bogenlänge benötigt dieses Verfahren allerdings keine Berechnung des Tangentenvektors  $y'(s_j)$  – weder im Prädiktor- noch im Korrektorschritt.

### Beispiel 40

Wir wenden den Algorithmus auf Beispiel 29 mit der Funktion

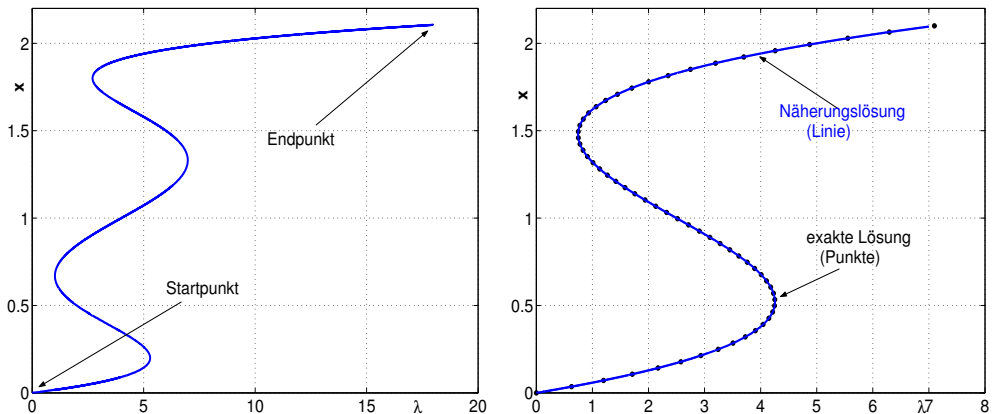
$$f(x, \lambda) := 40\mu x^5 - 200\mu x^4 + (342\mu + 8)x^3 - (226\mu + 24)x^2 + 0.5(91\mu + 37)x - \lambda$$

**ALGORITHMUS 39 (Gauß-Newton-Fortsetzung)**

Function  $[N, y] = \text{gaussnewton}(f, Df, a, b, x, \text{smax}, \text{tol})$

1. Setze  $j := 0$ ,  $\lambda_0 := a$ ,  $x^P := x$ ,  $s_0 := 0$   
Wähle Schrittweite  $h$  sowie Iterationszahl  $k_{\text{opt}}$ .
2. (*Startrechnung*) Bestimme  $x_0 = x(\lambda_0)$  mit Verfahren (33) oder (34).  
Setze  $y_0 = (x_0, \lambda_0)$ .
3. Do while (  $a \leq \lambda_j \leq b$  and  $s_j < \text{smax}$  )
  - 3.1. Setze  $j := j + 1$  und  $s_j := s_{j-1} + h$ .
  - 3.2. (*Sekanten-Prädiktor*) Bestimme  $v_0 := y^P$  mit Verfahren (57),  
falls  $j > 1$ , andernfalls mit Verfahren (56).
  - 3.3. (*Gauß-Newton-Korrektor*)
    - (a) Setze  $k := 0$ .
    - (b)  $A := Df(v_k) = f'(v_k)$
    - (c) Löse das reguläre lineare System
 
$$(AA^T) \cdot b = f(v_k), \quad b \in \mathbb{R}^n$$
    - (d)  $v_{k+1} := v_k - A^T b$ ,  $k := k + 1$
    - (e) Falls  $\|A^T b\|_2 > \text{tol}$ , so gehe zu Schritt (b).
    - (f) Setze  $y_j := v_k$ .
  - 3.4. (*Schrittweiten-Anpassung*) wie in Algorithmus 28
4. Return  $N := j$ ,  $y_j = (x_j, \lambda_j)$ ,  $j = 0(1)N$

an. Für festen Parameter  $\mu = 1$  erhalten wir mit Startwerten  $(x_0, \lambda_0) = (0, 0)$  und maximaler Fortsetzungsschrittweite  $h_{\text{max}} = 0.005$  in Abbildung 10 (links) die Lösung mit



**Abb. 10** Berechnete Lösung  $\mathcal{L}$  zu Beispiel 29 mit  $\mu = 1$  (links) und  $\mu = 0$  (rechts)

Genauigkeit  $tol = 10^{-3}$  über alle vier Umkehrpunkte hinweg (vgl. die exakte Darstellung in Abbildung 6). Auch bei  $\mu = 0$  arbeitet die Fortsetzung zuverlässig und liefert in Abbildung 10 (rechts) die Näherungslösung. Zum Vergleich wurden hier exakte Kurvenpunkte eingezeichnet. ■

### 3.4 Fortsetzung mehrdimensionaler Mannigfaltigkeiten

Die üblichen Anwendungsprobleme aus der Praxis hängen meist von mehreren Parametern  $\lambda_1, \lambda_2, \dots, \lambda_p$  ab. Um das Lösungsverhalten über einem Bereich des Parameter-raumes  $\mathbb{R}^p$  zu studieren, ist dann ein mehrdimensionales Fortsetzungsproblem zu lösen. Anstelle einer Lösungskurve ist nun eine *Lösungsmannigfaltigkeit* des  $n$ -dimensionalen Problems

$$f(y) = 0 \quad \text{mit } y = (x, \lambda) \in D \subset \mathbb{R}^{n+p}, \quad p \geq 1, \quad D \text{ offen} \quad (68)$$

mit Parametervektor  $\lambda \in \mathbb{R}^p$  zu finden. Die Idee der Kurvenverfolgung mittels Pseudo-Bogenlänge, einen Kurvenpunkt  $y_j$  durch die in Abbildung 6 (rechts) dargestellte Orthogonalitätsbedingung  $y'(s_{j-1})^T(y - y^P) = 0$  mit dem Tangentialvektor  $y'(s_{j-1})$  zu bestimmen, lässt sich auch auf mehrdimensionale Lösungsmannigfaltigkeiten verallgemeinern. Betrachten wir dazu das Problem (68) unter folgender

#### Voraussetzung 41

- i.  $f : D \subset \mathbb{R}^{n+p} \rightarrow \mathbb{R}^n$  sei  $C^r$ -glatt mit  $r \geq 2$  sowie
- ii. Die  $n \times (n+p)$ -Matrix  $f'(y) = Df(y)$  hat maximalen Rang  $n$  für alle  $y \in D$ .

Dann ist die Menge  $\mathcal{M} = \{y \in D \mid f(y) = 0\}$  eine  $p$ -dimensionale Mannigfaltigkeit in  $\mathbb{R}^{n+p}$  mit  $C^r$ -Glattheit (vgl. Rheinboldt (1994)). Zu einem Punkt  $y_j \in \mathcal{M}$  bilden

- $T_{y_j}\mathcal{M} = N(Df(y_j))$  den  $p$ -dimensionalen Tangentialraum von  $\mathcal{M}$  bei  $y_j$  und
- $N_{y_j}\mathcal{M} = \{N(Df(y_j))\}^\perp = R(Df(y_j)^T)$  den  $n$ -dimensionalen Normalraum,

der das orthogonale Komplement von  $T_{y_j}\mathcal{M}$  in  $\mathbb{R}^{n+p}$  ist. Nehmen wir nun an, wir hätten  $p$  paarweise orthonormierte Vektoren  $b_1, b_2, \dots, b_p \in \mathbb{R}^{n+p}$  zur Verfügung, so dass die damit erweiterte  $(n+p) \times (n+p)$ -Matrix

$$C := \begin{pmatrix} Df(y_j) \\ B^T \end{pmatrix} \quad \text{mit } B = (b_1, b_2, \dots, b_p) \in \mathbb{R}^{(n+p) \times p} \quad (69)$$

regulär ist. Dies ist während der Lösungsfortsetzung realisierbar, wenn beispielsweise  $\mathcal{B} = \{b_1, b_2, \dots, b_p\}$  eine orthonormale Basis des Tangentialraumes  $T_{y_{j-1}}\mathcal{M}$  eines vorhergehenden Lösungspunktes  $y_{j-1}$  darstellt.

Zur Bestimmung von  $y_j$  bilden wir in Analogie zu (61) am Prädiktorpunkt  $y^P$  das erweiterte System

$$g(y) := \begin{pmatrix} f(y) \\ B^T(y - y^P) \end{pmatrix} = 0, \quad g : D \subset \mathbb{R}^{n+p} \rightarrow \mathbb{R}^{n+p}, \quad (70)$$



indem wir die  $n$  Gleichungen um  $p$  Projektionsbedingungen  $b_i^T(y - y^P) = 0, i = 1(1)p$ , ergänzen. Wegen Voraussetzung 41(i) ist  $g(y)$   $C^2$ -glatt, und mit (69) ist die Jacobi-Matrix  $Dg(y_j)$  regulär. Damit sind Newton-Verfahren und vereinfachtes Newton-Verfahren aus Abschnitt 2 für  $g(y) = 0$  lokal in der Umgebung von  $y_j$  ausführbar und konvergieren Q-quadratisch bzw. Q-linear gegen eine gesuchte Lösung  $y_j$  mit  $g(y_j) = 0$ .

Nach Erhalt von  $y_j$  ist allerdings noch die Frage zu klären, wie eine Orthonormalbasis  $\mathcal{B}_{\text{neu}} = \{b_1, b_2, \dots, b_p\}$  von  $T = T_{y_j}\mathcal{M}$  für den nächsten Fortsetzungsschritt zu bestimmen ist. Dazu lösen wir mit der verfügbaren Matrix  $B$  die  $p$  linearen Gleichungssysteme

$$\begin{pmatrix} Df(y_j) \\ B^T \end{pmatrix} \tau = e_{n+i}, \quad i = 1(1)p, \quad (71)$$

deren rechte Seiten die  $(n + p)$ -dimensionalen Einheitsvektoren  $e_{n+i}$  sind.

#### Satz 42

Die Lösungen  $\tau_1, \tau_2, \dots, \tau_p$  von (71) sind eindeutig bestimmt und liefern eine Basis des Tangentialraumes  $T_{y_j}\mathcal{M}$ .

**Beweis:** Die Koeffizientenmatrix ist offenbar regulär. Aus den ersten  $n$  Gleichungen von (71) folgt  $Df(y_j)\tau_i = 0$  für  $i = 1(1)p$ , womit  $\tau_i \in T_{y_j}\mathcal{M}$  ist. Aus der zusammengefassten Darstellung  $B^T(\tau_1, \tau_2, \dots, \tau_p) = I_p$  der übrigen  $p$  Gleichungen mit der Einheitsmatrix  $I_p$  ergibt sich dann die lineare Unabhängigkeit der Vektoren  $\tau_1, \tau_2, \dots, \tau_p$ .  $\square$

Schließlich gewinnt man aus dieser Basis mittels *Gram-Schmidt-Orthogonalisierung* oder *QR-Zerlegung* eine Orthonormalbasis  $\mathcal{B}_{\text{neu}} = \{b_1, b_2, \dots, b_p\}$  des Tangentialraumes  $T = T_{y_j}\mathcal{M}$ . Mit diesen Vektoren definieren wir die Matrix  $B = (b_1, b_2, \dots, b_p) \in \mathbb{R}^{(n+p) \times p}$  und können damit alle Elemente  $y \in \mathbb{R}^{n+p}$  eindeutig in der Form

$$y = y_j + Bs + w \quad \text{mit } s \in \mathbb{R}^p \text{ und } w \in T^\perp$$

darstellen. Betrachten wir an einem derartigen Punkt  $y$  das erweiterte System für die Unbekannte  $w \in \mathbb{R}^{n+p}$

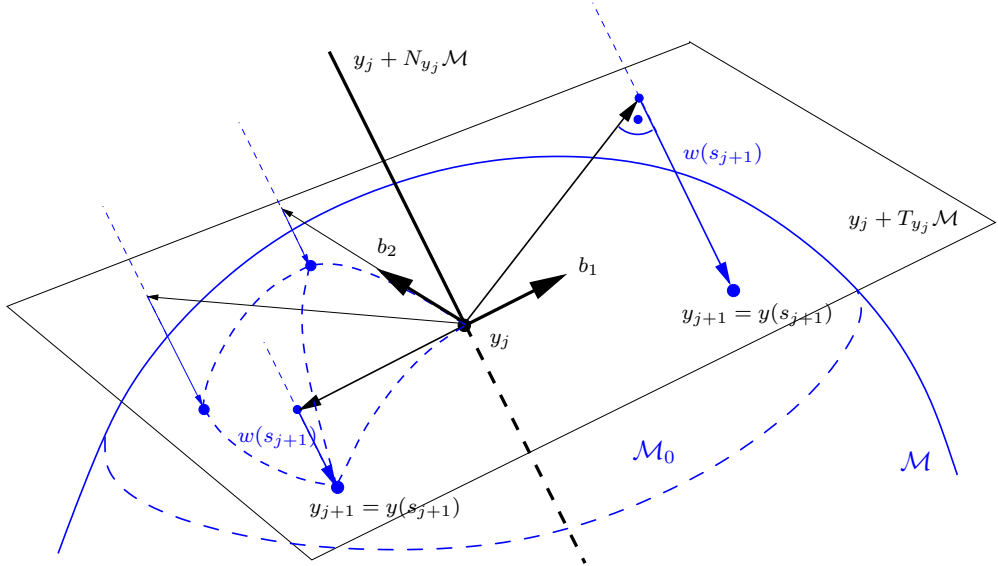
$$h(w, s) := \begin{pmatrix} f(y_j + Bs + w) \\ B^T w \end{pmatrix} = 0 \quad (72)$$

mit dem Parametervektor  $s \in \mathbb{R}^p$ . Die Abbildung  $h : \mathbb{R}^{n+p} \times \mathbb{R}^p \rightarrow \mathbb{R}^{n+p}$  hat offenbar dieselbe  $C^r$ -Glattheit wie  $f$ . Zudem gilt

$$h(0, 0) = 0 \quad \text{und} \quad D_w h(0, 0) = \begin{pmatrix} Df(y_j) \\ B^T \end{pmatrix} \text{ ist regulär.}$$

Nach dem impliziten Funktionentheorem existieren deshalb Umgebungen  $V_s(0) \subset \mathbb{R}^p$  und  $V_w(0) \subset \mathbb{R}^{n+p}$  von Null und eine eindeutige Abbildung  $w : V_s(0) \rightarrow V_w(0)$ , die der Gleichung  $h(w(s), s) = 0$  für alle  $s \in V_s(0)$  sowie  $w(0) = 0$  genügt. Damit kann die Mannigfaltigkeit  $\mathcal{M}$  an der Stelle  $y_j$  lokal durch die Funktion  $y : V_s(0) \rightarrow \mathbb{R}^{n+p}$  mit

$$y(s) := y_j + Bs + w(s), \quad s \in V_s(0), \quad w(0) = 0, \quad w(s) \in V_w(0) \subset T^\perp \quad (73)$$



**Abb. 11** Darstellung von  $y_j = y(s_j)$  und vier neuer Punkte  $y_{j+1} = y(s_{j+1})$  auf der Mannigfaltigkeit  $\mathcal{M}$

parametrisiert werden. Ausgehend vom Punkt  $y_j \in \mathcal{M}$  bestimmen wir durch geeignete Wahl von Parametervektoren  $s \in V_s(0)$  Nachbarnpunkte  $y(s)$  auf der Mannigfaltigkeit, indem wir System (72) nach  $w(s)$  auflösen. Für den dreidimensionalen Fall sind in Abbildung 11 die affinen Unterräume  $y_j + T_{y_j}\mathcal{M}$  mit der Orthonormalbasis  $\mathcal{B} = \{b_1, b_2\}$  sowie  $y_j + N_{y_j}\mathcal{M}$  mit Dimension 1 schwarz dargestellt. Ausgehend von  $y_j = y(s_j)$  werden z. B. vier neue Punkte  $y_{j+1} = y(s_{j+1})$  auf der blau dargestellten Mannigfaltigkeit  $\mathcal{M}$  erzeugt, indem zu vorgegebenen Parameterwerten  $s_{j+1}$  und damit berechneten Vektoren  $w(s_{j+1})$  (blau) die neuen Vektoren  $y_{j+1} := y_j + B s_{j+1} + w(s_{j+1})$  erzeugt werden. Auf der linken Bildseite erkennt man eine Triangulierung der Mannigfaltigkeit, die im praktischen Algorithmus geeignet vorzunehmen ist. Im nächsten Fortsetzungsschritt kann die Vorgehensweise für jeden Punkt  $y_{j+1}$  mit der zugehörigen Basis  $\mathcal{B}$  wiederholt werden.

#### Definition 44 (Moving Frame, begleitende Basis)

$\mathcal{M}_0$  sei eine offene Teilmenge von  $\mathcal{M}$ . Ein *Moving Frame* (begleitende Basis) der Klasse  $C^r$  besteht aus  $C^r$ -glatten Funktionen  $b_1(y), b_2(y), \dots, b_p(y)$ , die eine Basis  $\mathcal{B}$  des Tangentialraumes  $T_y\mathcal{M}$  für alle  $y \in \mathcal{M}_0$  bilden (vgl. Mei (2000)).

Wird eine feste Matrix  $B = (b_1, b_2, \dots, b_p) \in \mathbb{R}^{(n+p) \times p}$  so gewählt, dass die erweiterte Matrixfunktion

$$C(y) := \begin{pmatrix} Df(y) \\ B^T \end{pmatrix}$$

**ALGORITHMUS 43 (Fortsetzung von Mannigfaltigkeiten)**

Function  $[y_{\text{neu}}, B_{\text{neu}}] = \text{moving\_frame}(f, Df, y, s, B, \text{tol})$

**1. (Punkte-Bestimmung)**

For  $i = 1(1)i_{\max}$  do: // Durchlaufen von  $s = (s_1, \dots, s_{i_{\max}})$

**1.1. (Fortsetzungsparameter und Prädiktor)**

Setze  $s := s_i$  und Startpunkt  $w := 0$ .

**1.2. (Newton-Korrektor für (72)):**

Do while  $\|f(y + Bs + w)\| > \text{tol}$ :

$$\text{Löse } \begin{pmatrix} Df(y + Bs + w) \\ B^T \end{pmatrix} \Delta w = - \begin{pmatrix} f(y + Bs + w) \\ B^T w \end{pmatrix}$$

$$w := w + \Delta w$$

**1.3.  $y_{\text{neu}}^i := y + Bs + w$** **2. (Basis-Bestimmung)**

Mit  $y$  löse die linearen Systeme (74) nach  $\tau_1, \tau_2, \dots, \tau_p$ .

**3. (Orthonormierung)**

Orthonormiere  $\tau_1, \tau_2, \dots, \tau_p$ , womit die neue Orthonormalbasis

$B_{\text{neu}} = \{b_1, b_2, \dots, b_p\}$  entsteht.

**4. Return  $y_{\text{neu}}, B_{\text{neu}}$** 

für alle  $y \in \mathcal{M}_0$  regulär ist, so bilden die  $p$  Lösungen  $\tau_i(y)$  der Gleichungssysteme

$$C(y)\tau_i = e_{n+i}, \quad i = 1(1)p, \quad (74)$$

mit den  $(n + p)$ -dimensionalen Einheitsvektoren  $e_{n+i}$  einen Moving Frame für  $\mathcal{M}_0$ . Ein Prädiktor-Korrektor-Schritt dieser Vorgehensweise wird in Algorithmus 43 vorgestellt. Er berechnet zu gegebenem Punkt  $y_j$  und zugehöriger Orthonormalbasis  $B$  für die  $i_{\max}$  Parametervektoren  $s_1, \dots, s_{i_{\max}}$  die zugehörigen Punkte  $y_{\text{neu}}^i$ ,  $i = 1(1)i_{\max}$ , und aktualisiert die Orthonormalbasis in  $B_{\text{neu}}$ . Zur weiteren Umsetzung dieses Verfahrens, z. B. mit geeigneten Triangulierungen der Mannigfaltigkeit, empfehlen sich die Darstellungen von Henderson (2007) und Mei (2000).

### 3.5 Fortsetzung mit Ableitungsprädiktoren

In den nachfolgenden Abschnitten sind die Fortsetzungsprobleme für periodische und quasi-periodische Orbits durch folgende erschwerende Umstände charakterisiert:

- $f(x, \lambda)$  ist nicht durch einen exakten arithmetischen Ausdruck definiert, sondern Resultat eines numerischen Integrationsverfahrens für DGL. Die Jacobi-Matrix  $D_x f(x, \lambda)$  ist so nicht exakt verfügbar und das Newton-Verfahren nicht anwendbar.
- Der Zeitaufwand zur Berechnung eines einzigen Funktionswertes  $f(x, \lambda)$  beträgt ein Vielfaches des algebraischen Aufwandes zur Lösung des linearen Gleichungssystems im Korrektorschritt. Rechenzeitverhältnisse des „Funktionswertaufwandes“ zum rein algebraischen Aufwand von über 100 : 1 sind damit keine Seltenheit!
- Die Stabilitäts- und Bifurkationsanalyse erfordert zudem eine gute Approximation der Jacobi-Matrizen  $A^* := D_x f(x_j, \lambda_j)$ ,  $j = 0, 1, 2, \dots$ , um deren Spektren  $\sigma(A^*)$  genau ermitteln zu können.

Als Ausweg bietet sich der vereinfachte Newton-Korrektor (34) bei  $\lambda = \lambda_j$

$$u_{k+1} := u_k - (A^P)^{-1} f(u_k, \lambda_j), \quad k = 0, 1, 2, \dots \quad (75)$$

mit einer Differenzenapproximation  $A^P$  gemäß (35) an. Neben einem guten Prädiktorwert  $x^P = u_0$  mittels Prädiktoren hoher Ordnung ist eine genaue Approximation  $A^P$  der zu bestimmenden Jacobi-Matrix  $A_j := D_x f(x_j, \lambda_j)$  gefordert. Wenig geeignet sind dafür Matrizen  $A_{j-1}$  des vorhergehenden Fortsetzungsschrittes. Konstruiert man jedoch einen Prädiktor höherer Ordnung für  $A^P$  mit Hilfe der zurückliegenden  $A_j$ -Werte, so entstehen mit geringem Zusatzaufwand effiziente Fortsetzungsverfahren.

#### Beispiel 45

$A_j := D_x f(x_j, \lambda_j)$  bezeichne die Jacobi-Matrizen an den Punkten  $\lambda_j$  mit  $x_j = x(\lambda_j)$ . Dann konstruieren wir für die Ableitung an der Stelle  $\lambda_j$  den

- Basis-Prädiktor:  $A^P := A_{j-1}$ ,
- Sekanten-Prädiktor:  $A^P := A_{j-1} + \frac{\lambda_j - \lambda_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} \cdot [A_{j-1} - A_{j-2}]$  sowie die
- Polynom-Prädiktoren:  $A^P := \alpha_1 A_{j-1} + \alpha_2 A_{j-2} + \dots + \alpha_k A_{j-k}$ ,  $\alpha_i \in \mathbb{R}$ ,  
mittels polynomialer Extrapolation. ■

Um den modifizierten Fortsetzungsschritt in  $[\lambda_{j-1}, \lambda_j]$  zu verifizieren, benutzen wir einen Konvergenzsatz für approximative Newton-Verfahren von Kelley (1995). Wir betrachten dazu das nichtlineare System  $f(x, \lambda_j) = 0$  mit festem Parameterwert  $\lambda_j$ .

#### Satz 46

Falls  $x_j$  eine reguläre Nullstelle von  $f(x, \lambda_j)$  im Sinne der Definition 4 ist, so existieren Konstanten  $K_0 > 0$ ,  $\delta_1 > 0$  und  $\delta_2 > 0$ , so dass für alle  $u_0$  mit  $\|u_0 - x_j\| < \delta_1$  und für alle  $A \in \mathbb{R}^{n \times n}$  mit  $\|A - D_x f(x_j, \lambda_j)\| < \delta_2$  die Iteration  $u_{k+1} = u_k - A^{-1} f(u_k, \lambda_j)$ ,  $k = 0, 1, 2, \dots$ , Q-linear gegen  $x_j$  konvergiert. Mit  $e_k := u_k - x_j$  gilt die Abschätzung

$$\|e_{k+1}\| \leq K_0 (\|e_0\| + \|A - D_x f(x_j, \lambda_j)\|) \cdot \|e_k\|, \quad k = 0, 1, 2, \dots \quad (76)$$

Die Fehlerschätzung (76) zeigt eine gleiche Wichtung der Prädiktoren  $x^P$  und  $A^P$ , weshalb sich dieselbe Approximationsordnung für beide, d. h.  $x^P - x_j = O(h^p)$  und  $A^P - D_x f(x_j, \lambda_j) = O(h^p)$  empfiehlt. In Algorithmus 47 benutzen wir dazu den

#### ALGORITHMUS 47 (Fortsetzung mit Ableitungsprädiktor)

Function  $[N, \lambda, x] = \text{continuation2}(f, Df x, a, b, x, \text{tol})$

1. Setze  $j := 0$ ,  $\lambda_0 := a$ ,  $x^P := x$  und  $A^P := D_x f(x^P, \lambda_0)$ .  
Wähle Schrittweite  $h$  sowie Iterationszahl  $k_{\text{opt}}$ .
2. (Startrechnung) Bestimme  $x_0 = x(\lambda_0)$  mit Verfahren (75).  
Setze  $A_0 := D_x f(x_0, \lambda_0)$ .
3. Do while  $\lambda_j \leq b$ 
  - 3.1. Setze  $j := j + 1$ .
  - 3.2. Setze  $\lambda_j := \lambda_{j-1} + h$ . Falls  $\lambda_j > b$ , so  $\lambda_j := b$ .
  - 3.3. (Lösungsprädiktor)  
Bestimme  $x^P$  mit Hilfe von  $x_{j-1}, x_{j-2}, \dots, x_0$ , z. B. mit
$$x^P := x_{j-1} + \frac{\lambda_j - \lambda_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} \cdot [x_{j-1} - x_{j-2}].$$
  - 3.4. (Ableitungsprädiktor)  
Bestimme  $A^P$  mit Hilfe von  $A_{j-1}, A_{j-2}, \dots, A_0$ , z. B. mit
$$A^P := A_{j-1} + \frac{\lambda_j - \lambda_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} \cdot [A_{j-1} - A_{j-2}].$$
  - 3.5. (Korrektor) Bestimme  $x_j = x(\lambda_j)$  mit Verfahren (75).
  - 3.6. (Ableitungs-Bestimmung) Berechne  $A_j = D_x f(x_j, \lambda_j)$ .
  - 3.7. (Schrittweiten-Bestimmung) wie in Algorithmus 28
4. Return  $N := j$  sowie  $\lambda_j, x_j, j = 0(1)N$

bewährten Sekanten-Prädiktor. Um den Fortsetzungsalgorithmus zu verifizieren, führen wir entsprechend Formel (31) eine Approximationsordnung für beide Prädiktoren ein. Die Schrittweiten  $h_j$  seien darin beschränkt durch  $h_{\max} := \max_{j=1(1)N} h_j$ .

#### Voraussetzung 48

1. Der Lösungsprädiktor besitze die Ordnung  $p_1$ , d. h. es existiert ein  $K_1 > 0$ , so dass

$$\|x^P(\lambda) - x(\lambda)\| \leq K_1 h^{p_1}, \quad |h| < h_{\max} \quad (77)$$

für  $\lambda = \lambda_{j-1} + h$ , mit  $\lambda_{j-1}, \lambda \in \Lambda$ , gilt.

2. Der Ableitungsprädiktor besitze die Ordnung  $p_2$ , d. h. es existiert ein  $K_2 > 0$ , so dass

$$\|A^P(\lambda) - D_x f(x(\lambda), \lambda)\| \leq K_2 h^{p_2}, \quad |h| < h_{\max} \quad (78)$$

für  $\lambda = \lambda_{j-1} + h$ , mit  $\lambda_{j-1}, \lambda \in \Lambda$ , gilt.

Das folgende Lemma garantiert dann, dass das Korrektorverfahren für „hinreichend gute“ Prädiktorwerte  $x^P$  und  $A^P$  stets konvergiert:

**Lemma 49**

Unter den Voraussetzungen 23 und 48 existieren Konstanten  $K_0 > 0$ ,  $\delta_1 > 0$  und  $\delta_2 > 0$ , so dass folgende Behauptungen gelten:

- i.  $f(x, \lambda) = 0$  besitzt für alle  $\lambda \in \Lambda$  eine eindeutige Lösung  $x(\lambda)$ .
- ii. Zu festem  $\lambda_{j-1} \in \Lambda$  konvergiert das Korrektorverfahren mit  $\lambda_j = \lambda_{j-1} + h_j$

$$u_{k+1} = u_k - (A^P)^{-1} f(u_k, \lambda_j), \quad k = 0, 1, 2, \dots$$

gegen  $x_j = x(\lambda_j)$ , falls  $0 < h_j < H := \min \left\{ p_1 \sqrt{\delta_1 / K_1}, p_2 \sqrt{\delta_2 / K_2} \right\}$  gilt.

- iii. Falls  $H := \min_{i=1,2} \left\{ p_i \sqrt{\delta_i / K_i}, p_i \sqrt{1 / (4K_0 K_i)} \right\}$  gewählt wird, so gilt

$$\|e_{k+1}\| \leq K_0 (K_1 h^{p_1} + K_2 h^{p_2}) \|e_k\| \leq \frac{1}{2} \|e_k\|, \quad k = 0, 1, 2, \dots \quad (79)$$

für  $e_k = u_k - x(\lambda_j)$ .

Um Algorithmus 47 abzusichern, ist wie in Satz 27 eine explizite Auswahlregel für die Schrittweiten  $h_j$  anzugeben, die die Durchführbarkeit und Endlichkeit der Fortsetzung garantiert. Mittels vollständiger Induktion über  $j$  zeigt man leicht

**Satz 50**

Die Voraussetzungen 23 und 48 seien erfüllt. Ferner genügen die Startwerte  $x^P$  und  $A^P$  am Startpunkt  $\lambda_0 = a$  den Voraussetzungen

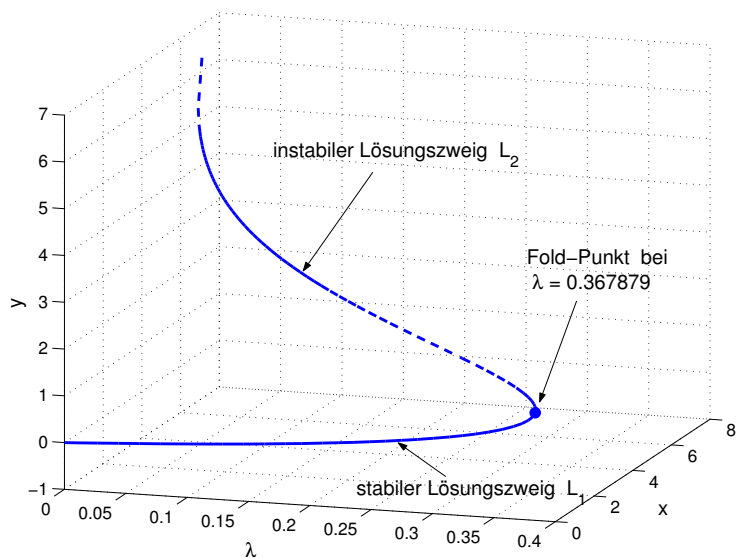
$$\|x^P - x(\lambda_0)\| < \delta_1 \quad \text{und} \quad \|A^P - D_x f(x(\lambda_0), \lambda_0)\| < \delta_2$$

mit den nach Lemma 49 garantierten Konstanten  $\delta_1 > 0$ ,  $\delta_2 > 0$ . Wählt man die Schrittweite  $h_j := \lambda_j - \lambda_{j-1}$  gemäß

$$h_j := \min \{H/2, b - \lambda_{j-1}\}, \quad \text{mit } H \text{ aus Teil iii. des Lemmas,} \quad (80)$$

so ist Algorithmus 47 durchführbar, d. h. er liefert zu jedem  $\lambda_j$  genau eine Lösung  $x(\lambda_j)$  der Gleichung  $f(x, \lambda_j) = 0$ . Er ist zudem endlich, d. h.  $\lambda_N = b$  für ein  $N \in \mathbb{N}$ .

Eine effiziente Wahl der Fortsetzungsschrittweiten  $h_j$  kann wie in Algorithmus 28 erfolgen. Die Berechnung der Jacobi-Matrizen  $A_j = D_x f(x_j, \lambda_j)$  in Schritt 3.6 sollte mit der Stabilitätsanalyse verbunden werden, die im nachfolgenden Abschnitt behandelt wird.



**Abb. 12** Lösungszweige  $L_1$  und  $L_2$  stabiler und instabiler Gleichgewichtslagen des Bratu-Gelfand-Systems (81)

**Beispiel 51**

Dhooge und Kuznetsov (2002) betrachten das Bratu-Gelfand-System mit Parameter  $\lambda$

$$\begin{aligned} \dot{x} &= f(x, y, \lambda) = y - 2x + \lambda e^x \\ \dot{y} &= f(y, x, \lambda) = x - 2y + \lambda e^y. \end{aligned} \tag{81}$$

Im Parameterintervall  $\lambda \in [0.01, 0.367879]$  besitzt das System zwei Gleichgewichtspunkte, die am kritischen Wert  $\lambda^* = 0.367879\dots$  zusammenfallen. Abbildung 12 zeigt die beiden Lösungszweige  $L_1$  (stabil) und  $L_2$  (instabil) sowie den Fold-Punkt  $(\lambda, x, y) = (0.367879\dots, 1, 1)$ . Wir notieren nun dieses System zu Testzwecken

$$\left. \begin{aligned} \dot{x}_i &= f_i(x, y, \lambda) = y_i - 2x_i + \lambda e^{x_i} \\ \dot{y}_i &= f_i(y, x, \lambda) = x_i - 2y_i + \lambda e^{y_i} \end{aligned} \right\} \quad \text{für } i = 1, 2, \dots, \frac{n}{2} \tag{82}$$

mit Systemgrößen  $n = 2, 20, 100, 200, 400$ , um damit folgende vier Versionen der Approximation der Jacobi-Matrizen zu vergleichen:

**Tab. 2** Anzahl N der Fortsetzungsschritte und Rechenzeit in sec (Pentium 4 M 1.60 GHz)

$n$	Version 1		Version 2		Version 3		Version 4	
	N	Zeit	N	Zeit	N	Zeit	N	Zeit
2	69	0.10	280	0.28	50	0.08	37	0.08
20	77	0.36	353	0.85	54	0.20	39	0.16
100	86	2.52	412	7.94	60	1.26	40	0.93
200	88	9.10	433	31.56	62	4.92	41	3.51
400	91	46.13	457	186.3	63	27.42	41	18.74

**Version 1:** Sekanten-Prädiktor  $x^P$  und übliche Approximation  $A^P := D_x f(x^P, \lambda_j)$

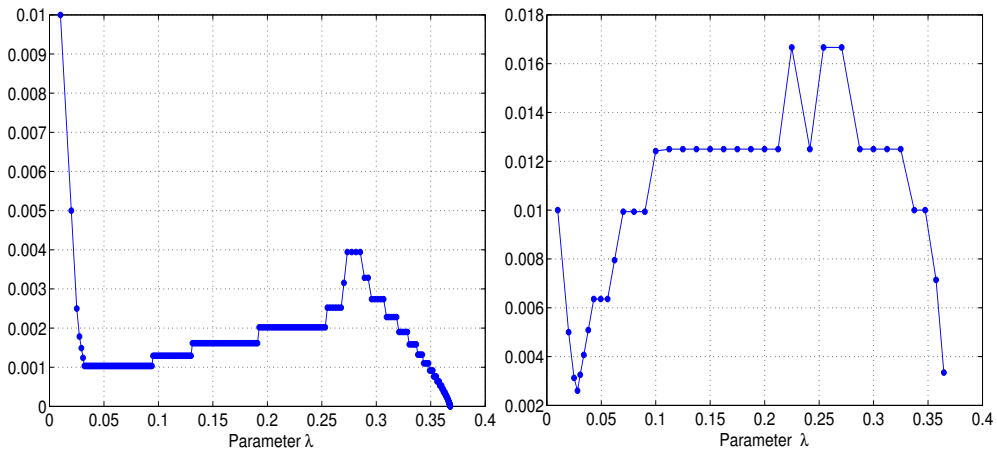
**Version 2:** Sekanten-Prädiktor  $x^P$  und Basis-Prädiktor  $A^P := A_{j-1}$

**Version 3:** Sekanten-Prädiktor  $x^P$  und Sekanten-Prädiktor  $A^P$  gemäß Algorithmus 47

**Version 4:** Prädiktor 2. Ordnung  $x^P$  und Prädiktor 2. Ordnung  $A^P$ .

Beginnend mit  $\lambda = 0.01$  verfolgen wir den instabilen Zweig  $L_2$  bis zu  $\lambda = 0.36787$  und analysieren bei jedem  $\lambda_j$  zudem die Stabilität. Startwerte sind  $x_i = y_i = 10$  für  $i = 1, 2, \dots, \frac{n}{2}$ ; die Genauigkeit ist stets  $\text{tol} = 10^{-10}$ .

Tabelle 2 zeigt die Überlegenheit der Versionen 3 und 4 mit gleicher Approximationsordnung von Lösung und Jacobi-Matrix. Version 4 kann so mit den größten internen Fortsetzungsschrittweiten arbeiten, wie man in Abbildung 13 leicht erkennt. ■



**Abb. 13** Interne Fortsetzungsschrittweiten  $h$  der Versionen 2 (links) und 4 (rechts)

Die Idee, Prädiktoren hoher Ordnung für die benötigten Jacobi-Matrizen zu benutzen, lässt sich mit entsprechendem Aufwand auch bei Bogenlängen-Parametrisierung und Gauß-Newton-Fortsetzung in den Algorithmen 34 und 39 leicht anwenden.

## 4 Stabilitäts- und Bifurkationsanalyse

Die Klärung des Stabilitätsverhaltens von Gleichgewichtslösungen des Systems

$$\dot{x} = f(x, \lambda), \quad f : D \times \mathbb{R} \rightarrow \mathbb{R}^n, \quad D \subset \mathbb{R}^n, \text{ offen}, \quad \lambda \in \Lambda = [a, b] \quad (83)$$

mit  $C^r$ -glattem Vektorfeld  $f \in C^r(D \times \mathbb{R})$  mit  $r \geq 1$ , ist von erheblicher praktischer Bedeutung, denn nur asymptotisch stabile Lösungen sind wegen unvermeidlicher Störungen in den Anfangswerten physikalisch relevant. Ein direkter Nachweis der Ljapunov-Stabilität mittels einer Ljapunov-Funktion ist im Allgemeinen nicht praktikabel, weshalb die indirekte Methode für das parameterabhängige System (83) eingesetzt wird.



## 4.1 Stabilitätsanalyse

Zu festem Parameterwert  $\lambda \in \Lambda$  sei  $x^0 = x^0(\lambda)$  ein Fixpunkt des Flusses. Wenn wir annehmen, dass die Lösung  $\varphi(\cdot; x)$  der DGL einen technischen Ablauf in einem Gerät beschreibt, dann ist es auf Grund von Messungenauigkeiten niemals möglich, beim erneuten Einschalten des Gerätes exakt den gleichen Ablauf zu erhalten (vgl. Hoffmann et al. (2006), S. 338). Vielmehr wird man stattdessen mit einem benachbarten (gestörten) Anfangszustand  $x$  starten. Um zu untersuchen, unter welchen Voraussetzungen die Prozessphasen  $\varphi(\cdot; x)$  und  $\varphi(\cdot; x^0)$  annähernd gleich verlaufen, bietet sich das Konzept der Ljapunov-Stabilität an.

### Definition 52 (Ljapunov-Stabilität eines Fixpunktes)

Es sei  $\varphi_t$  der Fluss der DGL (83), definiert für alle  $t \in \mathbb{R}$  und festen Parameter  $\lambda$ .

- i. Ein Fixpunkt  $x^0$  heißt *Ljapunov-stabil*, falls zu jedem  $\varepsilon > 0$  eine Zahl  $\delta > 0$  existiert, so dass für alle  $x$  mit  $\|x - x^0\| < \delta$  und  $t \geq 0$  stets  $\|\varphi_t(x) - x^0\| < \varepsilon$  gilt.
- ii. Ein Fixpunkt  $x^0$  heißt *asymptotisch Ljapunov-stabil*, falls er stabil ist und wenn eine Konstante  $\alpha > 0$  existiert, so dass für alle  $x$  mit  $\|x - x^0\| < \alpha$  folgt  $\lim_{t \rightarrow \infty} \|\varphi_t(x) - x^0\| = 0$ .
- iii. Ein Fixpunkt  $x^0$ , der nicht stabil ist, heißt *instabil*.

Die Definition der Ljapunov-Stabilität besagt: Jede Trajektorie, die in  $x$  startet, kann die Kugel  $K(x^0; \delta)$  mit Radius  $\delta > 0$  verlassen, muss jedoch in der Kugel  $K(x^0; \varepsilon)$  mit Radius  $\varepsilon > 0$  bleiben. Im Unterschied zur asymptotischen Stabilität wird oft auch mit dem Begriff *Attraktivität* gearbeitet: Zu jedem Anfangswert  $x$  mit  $\|x - x^0\| < \eta$  existiert eine Lösung  $\varphi_t$  für alle  $t > 0$  und es gilt  $\lim_{t \rightarrow \infty} \|\varphi_t(x) - x^0\| = 0$ . Die asymptotische Stabilität bedeutet damit nichts anderes als Stabilität plus Attraktivität.

Definition 52 bietet allerdings kein praktikables Kriterium, wie man anhand des Vektorfeldes  $f$  die Stabilität eines Fixpunktes erkennt. Aus dem Satz von Hartman und Grobman lassen sich mit Hilfe der Linearisierung von  $f$  bei  $x^0$  Aussagen über die Stabilität einer Gleichgewichtslage  $x^0$  des DGL-Systems (83) gewinnen.

Zu festem Parameterwert  $\lambda \in \Lambda$  sei nun  $x^0 = x^0(\lambda)$  ein hyperbolischer Fixpunkt des Flusses. Bestimmen wir die Jacobi-Matrix  $A := A(\lambda) = D_x f(x^0(\lambda), \lambda)$  und deren Eigenwerte  $\sigma(A(\lambda)) = \{\mu_1(\lambda), \mu_2(\lambda), \dots, \mu_n(\lambda)\}$ , die sogenannten *charakteristischen Exponenten*, so liefert uns das Stabilitätskriterium die folgende numerisch leicht überprüfbare Aussage:

**Satz 53 (Stabilitätskriterium für nichtlineare DGL-Systeme)**

$x^0$  sei ein hyperbolischer Fixpunkt mit Jacobi-Matrix  $A := A(\lambda) = D_x f(x^0(\lambda), \lambda)$  und den charakteristischen Exponenten  $\mu_1(\lambda), \mu_2(\lambda), \dots, \mu_n(\lambda)$ .

- Ist der Realteil  $\operatorname{Re} \mu_i(\lambda) < 0$  für alle  $i \in \{1, 2, \dots, n\}$ , so ist  $x^0(\lambda)$  asymptotisch stabil.
- Existiert ein  $i \in \{1, 2, \dots, n\}$  mit  $\operatorname{Re} \mu_i(\lambda) > 0$ , so ist  $x^0(\lambda)$  instabil.

**Beispiel 54**

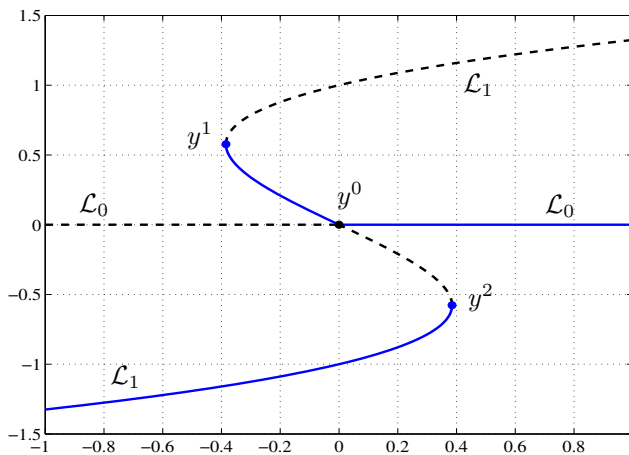
Betrachten wir das eindimensionale Beispiel mit dem Systemparameter  $\lambda \in [-1, 1]$

$$\dot{x} = f(x, \lambda) = x(x^3 - x - \lambda),$$

das die beiden Zweige  $\mathcal{L}_0 = \{(x, \lambda) \mid x = 0\}$  und  $\mathcal{L}_1 = \{(x, \lambda) \mid x^3 - x - \lambda = 0\}$  von Gleichgewichtslösungen besitzt. Auf  $\mathcal{L}_1$  berechnen wir die Ableitung

$$D_x f(x, \lambda) = 4x^3 - 2x - \lambda = 3x(x^2 - 1/3),$$

woraus sich der Eigenwert  $\mu = 3x(x - \sqrt{3}/3)(x + \sqrt{3}/3)$  ergibt. Mit  $\mu = 0$  erhalten wir die kritischen Punkte  $y^0 = (x^0, \lambda_0) = (0, 0)$ ,  $y^1 = (x^1, \lambda_1) = (\sqrt{3}/3, -2\sqrt{3}/9)$  und  $y^2 = (x^2, \lambda_2) = (-\sqrt{3}/3, 2\sqrt{3}/9)$  in Abbildung 14. Ein Stabilitätswechsel erfolgt offenbar nur an diesen drei Punkten, die die beiden Lösungszweige  $\mathcal{L}_0$  und  $\mathcal{L}_1$  in die dargestellten Abschnitte mit asymptotisch stabilem bzw instabilem Verhalten zerlegen.



**Abb. 14** Lösungszweige  $\mathcal{L}_0$  und  $\mathcal{L}_1$  zu Beispiel 54: stabile Zweige (blau) und instabile Zweige (schwarz gestrichelt)

Wir modifizieren die DGL nun geringfügig zu der Gleichung

$$\dot{x} = f(x, \lambda) = x(x^3 - x - \lambda^2), \quad \lambda \in [-1, 1],$$

und betrachten deren trivialen Lösungszweig  $\mathcal{L}_0 = \{(x, \lambda) \mid x = 0\}$ . Wegen der Ableitung  $D_x f(x, \lambda) = 4x^3 - 2x - \lambda^2$  lautet der Eigenwert  $\mu = -\lambda^2 < 0$  für alle  $\lambda \neq 0$ . Also ist die Lösung  $x^0 = 0$  asymptotisch stabil für alle  $\lambda \neq 0$ . Bei  $\lambda_0 = 0$  verschwindet zwar der Realteil  $\operatorname{Re} \mu$ , es findet jedoch kein Stabilitätswechsel statt. ■

Über nicht-hyperbolische Fixpunkte kann im nichtlinearen Fall mit dem Eigenwertkriterium nicht entschieden werden. Ist dagegen die *Spektralabszisse* der Matrix  $A$

$$\nu(A) := \max_{\mu_i \in \sigma(A)} \operatorname{Re} \mu_i \quad (84)$$

negativ, so ist  $x(t) = x^0$  asymptotisch stabile Gleichgewichtslage. Damit können wir leicht den Schritt 3.6 in Algorithmus 47 modifizieren:

#### ■ Stabilitätsanalyse

- (a) Berechne die Jacobi-Matrix  $A_j := D_x f(x^j, \lambda_j)$ .
- (b) Bestimme die Eigenwerte  $\mu_1, \mu_2, \dots, \mu_n$  von  $A_j$   
und die Spektralabszisse  $\nu := \nu(A_j)$ .
- (c) Falls  $\nu < 0$ , so “ASYMPTOTISCH STABIL“,  
sonst falls  $\nu > 0$ , so “INSTABIL“,  
sonst “KEINE AUSSAGE MÖGLICH“.

#### Beispiel 55

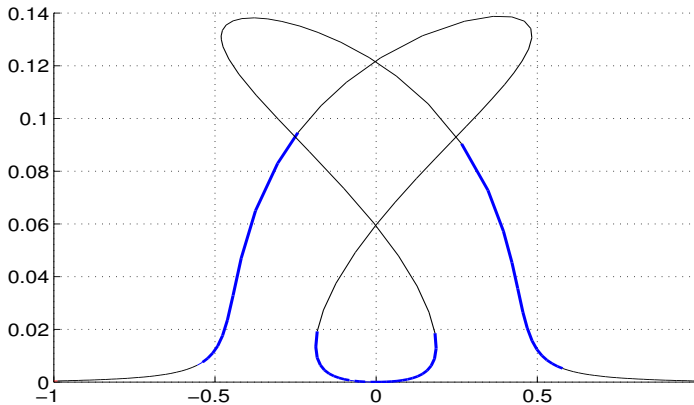
Wir betrachten nochmals Beispiel 35 mit  $f(x, \lambda)$ ,  $x = (x_1, x_2, x_3, x_4, x_5)^T$ , gegeben durch die fünf dortigen Funktionen. Die Steuervariable ist  $u = 0$  und der Parameterbereich  $-1 \leq \lambda \leq 1$ . Abbildung 15 zeigt die numerisch berechnete Lösung  $x_4$ , dargestellt über dem Parameter  $\lambda$ , wobei nun die Punkte mit nachgewiesener Instabilität schwarz gezeichnet wurden. Offenbar sind die Stabilitätswechsel noch genauer einzugrenzen. ■

## 4.2 Detektierung lokaler Bifurkationen

Mit der Stabilitätsanalyse kann das Auftreten lokaler Bifurkationen während der Lösungsfortsetzung entdeckt (detektiert) werden. Bei Marx und Vogt (2010) werden in Abschnitt 2.7.1 zur Jacobi-Matrix  $A := A(\lambda) = D_x f(x^0(\lambda), \lambda)$  der

- stabile Unterraum  $E^s$  der Dimension  $n_s = \dim E^s =$  Zahl der  $\mu_i$  mit  $\operatorname{Re} \mu_i < 0$
- instabile Unterraum  $E^u$  der Dimension  $n_u = \dim E^u =$  Zahl der  $\mu_i$  mit  $\operatorname{Re} \mu_i > 0$
- Zentrumsunterraum  $E^c$  der Dimension  $n_c = \dim E^c =$  Zahl der  $\mu_i$  mit  $\operatorname{Re} \mu_i = 0$

mit  $n_s + n_u + n_c = n$  definiert. Bestimmen wir im  $j$ -ten Fortsetzungsschritt diese Dimensionen  $n_s^j, n_u^j$  und  $n_c^j$ , so signalisiert eine Änderung dieser Werte bei zwei aufeinander folgenden Schritten die Möglichkeit einer lokalen Bifurkation im (hinreichend kleinen)



**Abb. 15**  
Lösungskomponente  $x_4$  über dem Parameter  $\lambda$ : stabile Zweige (blau) und instabile Zweige (schwarz)

Intervall  $[\lambda_{j-1}, \lambda_j]$ . Damit kann Schritt 3.6 in Algorithmus 47 leicht modifiziert werden:

#### ■ Bifurkationsdetektierung

- (a) Berechne die Jacobi-Matrix  $A_j := D_x f(x^j, \lambda_j)$ .
- (b) Bestimme die Eigenwerte  $\mu_1, \mu_2, \dots, \mu_n$  von  $A_j$  und die Dimensionen  $n_s^j, n_u^j, n_c^j$  der Unterräume  $E^s, E^u, E^c$ .
- (c) Falls  $n_s^j \neq n_s^{j-1} \vee n_u^j \neq n_u^{j-1} \vee n_c^j \neq n_c^{j-1}$ , so  
"BIFURKATION IN  $[\lambda_{j-1}, \lambda_j]$  MÖGLICH".

Die ständige Berechnung aller charakteristischen Exponenten  $\mu_i$  während der Lösungsfortsetzung ist bei großen Systemen sehr zeitintensiv. So gewinnen einfachere Kriterien an Bedeutung. Allerdings sind diese im Allgemeinen nur notwendig, nicht aber hinreichend für den jeweiligen Bifurkationstyp. Der Nachweis hinreichender Bedingungen ist meistens nicht praktikabel.

Wir nehmen desweiteren an, dass einer der folgenden Bifurkationstypen auftreten kann:

- Sattel-Knoten-Bifurkation (desweiteren auch als Falten-Bifurkation bezeichnet)
- transkritische Bifurkation
- Heugabel-Bifurkation (Pitchfork-Bifurkation) sowie
- Hopf-Bifurkation.

Mit der zusammenfassenden Bezeichnung  $y = (x, \lambda)$  und  $f(y) = f(x, \lambda)$  sei  $y^0 = (x^0, \lambda_0)$  mit dem Gleichgewichtspunkt  $x^0$  gegeben, d. h.  $f(y^0) = 0$ . Am Punkt  $y^0$  definieren wir die Matrizen der Dimension  $n \times n$  und  $n \times (n+1)$

$$A := D_x f(x^0, \lambda_0) \quad \text{und} \quad B := f'(y^0) = (D_x f(x^0, \lambda_0), D_\lambda f(x^0, \lambda_0)).$$

Über die Nullräume von  $A$  und  $B$  gestattet die Theorie dann folgende Aussage:

**Satz 56**

Es sei  $f \in C^r(D \times \Lambda)$ ,  $r \geq 2$ . Dann sind die Nullraumdimensionen

$$\dim N(A) = \begin{cases} 0, & \text{falls } x^0 \text{ regulärer oder Hopf-Punkt} \\ 1, & \text{falls } x^0 \text{ Falten-Punkt, transkritischer oder Heugabel-Punkt ist.} \end{cases}$$

$$\dim N(B) = \begin{cases} 1, & \text{falls } x^0 \text{ regulärer Punkt} \\ 1, & \text{falls } x^0 \text{ Falten- oder Hopf-Punkt} \\ 2, & \text{falls } x^0 \text{ transkritischer oder Heugabel-Punkt ist.} \end{cases}$$

Bei Parametrisierung der Lösungskurve  $y = y(s) = (x(s), \lambda(s))$  nach ihrer Bogenlänge oder Pseudo-Bogenlänge  $s$  steht zudem der normierte Tangentenvektor  $y'(s_0)$  am betrachteten Punkt  $y^0 = (x^0, \lambda_0) = (x(s_0), \lambda(s_0))$  zur Verfügung. Damit definieren wir die erweiterte Matrix

$$C := \begin{pmatrix} f'(y(s_0)) \\ y'(s_0)^T \end{pmatrix} = \begin{pmatrix} B \\ y'(s_0)^T \end{pmatrix}, \quad C \in \mathbb{R}^{(n+1) \times (n+1)}.$$

Wegen  $f(y(s)) = 0$  folgt durch Differentiation  $B \cdot y'(s_0) = 0$ , so dass  $y'(s_0)$  linear unabhängig von den Zeilen von  $B$  ist. Damit ist die Nullraumdimension von  $C$  um Eins kleiner als diejenige von  $B$ . Mit den Matrizen  $A, B$  und  $C$  ist dann die in Tabelle 3 angegebene teilweise Klassifikation der Bifurkationspunkte möglich.

**Tab. 3** Nullraumdimension zur Klassifikation der Punkte  $x^0$

Typ der Bifurkations- punkte $x^0$	Parametrisierung nach $\lambda$		Parametrisierung nach $s$
	$A$	$B$	$C$
regulärer Punkt	0	1	0
Falten-Bifurkation	1	1	0
transkritische Bifurkation	1	2	1
Heugabel-Bifurkation	1	2	1
Hopf-Bifurkation	0	1	0

Um die numerisch wenig geeigneten Dimensions- und Rangbestimmungen zu vermeiden, führen wir mit Hilfe der Determinanten von  $A$  und  $C$  skalare Testfunktionen  $\psi$  ein. Mit  $C = (x(s), \lambda(s))$  bezeichnen wir den verfolgten Zweig von Gleichgewichtslösungen mit Parameter  $s \in \mathbb{R}$ .

**Definition 57 (Testfunktion)**

Es sei  $\mathcal{U}$  eine Umgebung von  $C$ . Eine stetige Funktion  $\psi : \mathcal{U} \rightarrow \mathbb{R}$ , die beim Durchlaufen der Kurve  $C = (x(s), \lambda(s))$  genau in  $s = s_0$  ihr Vorzeichen (strikt) ändert, heißt *Testfunktion* für den Bifurkationspunkt  $(x^0, \lambda_0) = (x(s_0), \lambda(s_0))$ .

Ist die Testfunktion stetig differenzierbar, so garantieren die Eigenschaften

$$\psi(x(s_0), \lambda(s_0)) = 0, \quad \left. \frac{d}{ds} \psi(x(s), \lambda(s)) \right|_{s=s_0} \neq 0, \quad (85)$$

dass  $s_0$  eine reguläre Nullstelle ist und der Bifurkationspunkt  $(x^0, \lambda_0)$  direkt während der Lösungsfortsetzung beobachtet werden kann. Wir wollen nun geeignete Testfunktionen für die genannten Bifurkationstypen einführen.

**Detektierung der Sattel-Knoten-Bifurkation (Falten-Bifurkation)**

Definieren wir mit den Determinanten die beiden Funktionen auf dem Lösungszweig  $C$

$$\psi_F(y) := \det(D_x f(y)) \quad \text{und} \quad \psi_B(y) := \det \begin{pmatrix} f'(y) \\ y'(s)^T \end{pmatrix}, \quad (86)$$

so können  $\psi_F(x(s), \lambda(s))$  und  $\psi_B(x(s), \lambda(s))$  als Testfunktionen fungieren. Aus Tabelle 3 und der Eigenwertdarstellung  $\psi_F(y) = \det(D_x f(y)) = \mu_1 \mu_2 \cdots \mu_n$  gewinnen wir zusammen mit den Eigenschaften der Bifurkationstypen aus Marx und Vogt (2010)

**Satz 58**

Es sei  $f \in C^r(D \times \Lambda)$ ,  $r \geq 2$ .  $y^0 = (x^0, \lambda_0)$  gegeben mit  $f(y^0) = 0$ .

- i.  $\psi_F(y)$  und  $\psi_B(y)$  sind stetig differenzierbar.
- ii. Ist  $x^0$  regulärer oder Hopf-Punkt, so ist  $\psi_F(y^0) \neq 0$ .
- iii. Liegt bei  $x^0$  eine Falten-Bifurkation vor, so ist  $\psi_F(y^0) = 0$  und  $\psi_B(y^0) \neq 0$ .
- iv. Ist  $x^0$  ein Falten-Punkt, so wechselt  $\psi_F(y)$  bei  $y = y^0$  das Vorzeichen.

**Beispiel 59**

In Beispiel 54 mit der Funktion  $f(x, \lambda) = x(x^3 - x - \lambda)$  verfolgen wir nun den Lösungszweig  $\mathcal{L}_1 = \{(x, \lambda) \mid x^3 - x - \lambda = 0\}$  aus Abbildung 14. In der Umgebung des Kurvenpunktes  $y^1 = (x^1, \lambda_1) = (\sqrt{3}/3, -2\sqrt{3}/9)$  parametrisieren wir die Kurve nach  $x$ , d. h.  $y(s) = (x(s), \lambda(s)) = (s, s^3 - s)$ , und erhalten damit den Wert der Testfunktion

$$\psi_F(y(s)) = 4x(s)^3 - 2x(s) - \lambda(s) = 3s^3 - s, \quad \text{also} \quad \psi_F(y^1) = 0.$$

Wegen der Ableitung an dieser Stelle  $y^1 = y(s_1)$  mit  $s_1 = \sqrt{3}/3$

$$\left. \frac{d}{ds} \psi(y(s)) \right|_{s=s_1} = 9s_1^2 - 1 = 2 > 0$$

erfolgt bei  $y^1$  ein Nulldurchgang. Für den Tangentenvektor am Punkt  $y^1$  erhalten wir  $y'(s_1) = (x'(s_1), \lambda'(s_1))^T = (1, 0)^T$ , womit wir die erweiterte Matrix

$$C = \begin{pmatrix} f'(y(s_1)) \\ y'(s_1)^T \end{pmatrix} = \begin{pmatrix} 0 & -\sqrt{3}/3 \\ 1 & 0 \end{pmatrix}$$

berechnen und für die zweite Testfunktion  $\psi_B(y_1) = \det C = \sqrt{3}/3 \neq 0$  erhalten. ■

### Detektierung der transkritischen und Heugabel-Bifurkation

Falls  $\psi_F(y^0) = 0$  und  $\psi_B(y^0) = 0$  wird, so ist ein Verzweigungspunkt im engeren Sinne (branching point) – in der Form einer transkritischen bzw. Heugabel-Bifurkation – zu vermuten. Dieses Kriterium ist zwar notwendig, jedoch nicht hinreichend. Die Überprüfung hinreichender Bedingungen ist in praktischen Fällen meist zu aufwändig, weshalb man es bei der Berechnung von  $\psi_F(y^0)$  und  $\psi_B(y^0)$  belässt.

### Beispiel 60

Wir setzen Beispiel 59 fort und untersuchen den Punkt  $y^0 = (x^0, \lambda_0) = (0, 0)$ . Offenbar ist  $\psi_F(y^0) = 4x(s)^3 - 2x(s) - \lambda(s) = 3s^3 - s$ , also  $\psi_F(y^0) = 0$ . Wir berechnen nun  $\psi_B(y^0)$  auf beiden Lösungszweigen: Auf  $\mathcal{L}_0 = \{(x, \lambda) \mid x = 0\}$  ist  $f'(0, 0) = (0, 0)$  und der Tangentenvektor trivialerweise  $y'(0) = (0, 1)^T$ , womit  $\psi_B(y^0) = \det C = 0$  folgt. Auch auf dem zweiten Lösungszweig  $\mathcal{L}_1$  erhalten wir für  $y^0 = 0$  den Wert  $\psi_B(y^0) = 0$ , womit bei der numerischen Verfolgung beider Zweige stets ein Verzweigungspunkt im engeren Sinne – in diesem Falle mit transkritischer Bifurkation – signalisiert wird. ■

Ein wesentliches Problem beim Auftreten von Verzweigungspunkten  $y^* = (x^*, \lambda^*)$  stellt deren Überwindung während der Parameterfortsetzung dar. Denn anders als bei Sattel-Knoten-Bifurkationen (Falten-Punkten) ist nun sowohl die Jacobi-Matrix  $A = D_x f(y^*)$  als auch die für die Bogenlängen-Parameterisierung oder die Gauß-Newton-Fortsetzung wesentliche  $(n + 1) \times (n + 1)$ -Matrix  $C$  singulär und damit die grundlegende Voraussetzung 30 iii dieser beiden Methoden verletzt. Damit versagt in der Regel der Newton-Korrektor aller bisherigen Fortsetzungsalgorithmen am Bifurkationspunkt  $y^*$ . Der Einzugsbereich des Korrektorverfahrens schrumpft bei Annäherung der Kurvenpunkte  $y(s)$  gegen  $y^*$  auf Null; das betrifft genauso den anderen schneidenden Lösungszweig. Allerdings beweisen Allgower und Georg (1990) für die Gauß-Newton-Fortsetzung 39, dass eine in Abbildung 16 dargestellte Kegelumgebung  $K \subset \mathbb{R}^{n+1}$  existiert, so dass das Korrektorverfahren für alle Startwerte  $y^0 \in K$  konvergiert:

### Satz 61

Es sei  $y^* \in \mathbb{R}^{n+1}$  transkritischer oder Heugabel-Bifurkationspunkt. Dann existieren eine offene Umgebung  $\mathcal{U}$  von  $\mathcal{L}_{reg} = \{y \mid f(y) = 0, y \text{ regulärer Punkt von } \mathcal{L}\}$  und Konstanten  $\varepsilon > 0$ ,  $\delta > 0$ , so dass gilt:

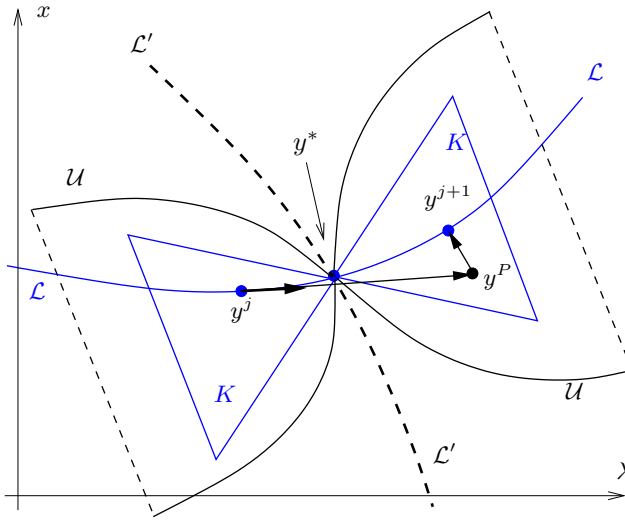
- i. Der Gauß-Newton-Korrektor in Algorithmus 39 konvergiert für jeden Startpunkt  $y^0 \in \mathcal{U}$  gegen einen Punkt  $y^\infty \in \mathcal{L}$ .

ii. Die Abschätzungen des Satzes 38 bleiben für  $y^0, y^1, y^\infty$  gültig.

iii. In der Umgebung  $\mathcal{U}$  existiert ein Kegel  $K \subset \mathcal{U}$  mit

$$K = \{y \mid y = y^* + [y'(s_0) + z]s\} \quad \text{für } 0 < |s| < \varepsilon \quad \text{und} \quad \|z\|_2 < \delta.$$

In Abbildung 16 sind die sich in  $y^*$  schneidenden Lösungswege  $\mathcal{L}$  und  $\mathcal{L}'$  mit der Umgebung  $\mathcal{U}$  und dem Kegel  $K$  dargestellt. Um vom letzten bestimmten Kurvenpunkt  $y^j$



**Abb. 16** Darstellung der sich schneidenden Lösungswege  $\mathcal{L}$  und  $\mathcal{L}'$  in der Umgebung des Verzweigungspunktes  $y^*$  sowie der Umgebung  $\mathcal{U}$  und des Kegels  $K$

während der Fortsetzung stets in der Kegelmenge zu verbleiben, wird ein „Überspringen“ des Bifurkationspunktes  $y^*$  empfohlen: Mit hinreichend großer Fortsetzungsschrittweite ist ein Prädiktorwert  $y^P$  für den nächsten Kurvenpunkt  $y^{j+1}$  so zu bestimmen, dass er wiederum im Kegel  $K$  liegt. Dafür empfehlen sich besonders die in Abschnitt 3.1 angeführten Prädiktoren hoher Ordnung unter Benutzung zurückliegender Kurvenpunkte. Die algorithmische Umsetzung überlassen wir jedoch dem interessierten Leser.

### Detektierung der Hopf-Bifurkation

Da im Fall eines Hopf-Punktes  $y^*$  die Jacobi-Matrix  $A = D_x f(y^*)$  regulär ist, sind die in (86) eingeführten  $\psi_F(y)$  und  $\psi_B(y)$  nicht als Testfunktionen zu gebrauchen. Das skiz-



zierte Vorgehen zur Bifurkationsdetektierung mittels der Unterräume  $E^s, E^u, E^c$  kann durch folgende skalare Testfunktion vereinfacht werden

$$\begin{aligned}\psi_H(y) &:= \prod_{i>k} [\mu_i(y) + \mu_k(y)] \\ &= (\mu_2(y) + \mu_1(y))(\mu_3(y) + \mu_1(y)) \cdots (\mu_n(y) + \mu_1(y)) \cdot \\ &\quad \cdot (\mu_3(y) + \mu_2(y)) \cdots (\mu_n(y) + \mu_2(y)) \cdots (\mu_n(y) + \mu_{n-1}(y)).\end{aligned}\tag{87}$$

Führen wir einige Eigenschaften von  $\psi_H$  an, um sie als Testfunktion einsetzen zu können:

**Satz 62**

Es sei  $f \in C^r(D \times \Lambda)$ ,  $r \geq 2$ .  $y^* = (x^*, \lambda^*)$  gegeben mit  $f(y^*) = 0$ .

- i.  $\psi_H : D \times \Lambda \rightarrow \mathbb{R}$  ist eine glatte reellwertige Funktion.
- ii. Falls  $y^*$  Hopf-Punkt ist, so  $\psi_H(y^*) = 0$ .
- iii. Bei Parameterfortsetzung nach  $\lambda$  wechselt  $\text{Re } \psi_H(y)$  für  $\lambda = \lambda^*$  das Vorzeichen.

Eigenschaft (ii) folgt ohne Beschränkung der Allgemeinheit für einen Hopf-Punkt aus  $\mu_2(y^*) = \mu_1(y^*)$  mit  $\text{Re } \mu_1(y^*) = 0$  wegen  $\mu_1(y^*) + \mu_2(y^*) = 2\text{Re } \mu_1(y^*) = 0$ . Vermittels der Transversalitätsbedingung des Satzes von Hopf zeigt man unmittelbar, dass  $\frac{d}{d\lambda} \text{Re } \psi_H(x(\lambda), \lambda) \neq 0$  für  $\lambda = \lambda^*$  gilt. Ist  $y^*$  ein Falten-, Heugabel- oder transkritischer Bifurkationspunkt mit  $\mu_1(y^*) = 0$ , so ist zusätzlich zu fordern, dass  $\mu_i(y^*) \neq -\mu_k(y^*)$  für alle  $i, k > 1$  ist, um  $\psi_H(y^*) \neq 0$  zu garantieren. Eine Darstellung dieser Testfunktion, die gänzlich ohne Eigenwertbestimmung oder -approximation auskommt, scheint schwer möglich. Wir führen dazu wie Kuznetsov (1995) das sogenannte bialternierende Produkt  $A \odot B$  zweier Matrizen ein:

**Definition 63 (Bialternierendes Produkt)**

Das *bialternierende Produkt*  $C = A \odot B$  zweier Matrizen  $A = (a_{ik}) \in \mathbb{R}^{n \times n}$  und  $B = (b_{ik}) \in \mathbb{R}^{n \times n}$  ist eine  $m \times m$ -Matrix mit  $m = n(n-1)/2$ . Die Zeilen von  $C$  werden mit dem Multiindex  $(p, q)$ ,  $p = 2(1)n$ ,  $q = 1(1)p - 1$ , die Spalten mit dem Multiindex  $(r, s)$ ,  $r = 2(1)n$ ,  $s = 1(1)r - 1$  bezeichnet. Die Matrixelemente berechnet man gemäß

$$C_{(p,q),(r,s)} := \frac{1}{2} \left\{ \begin{vmatrix} a_{pr} & a_{ps} \\ b_{qr} & b_{qs} \end{vmatrix} + \begin{vmatrix} b_{pr} & b_{ps} \\ a_{qr} & a_{qs} \end{vmatrix} \right\}.\tag{88}$$

Die Multiindizes  $(p, q)$  und  $(r, s)$  sind beispielsweise im Falle  $n = 4$  folgendermaßen angeordnet:

$$\begin{array}{cccc}
(2, 1), (2, 1) & (2, 1)(3, 1) & \dots & (2, 1), (4, 3) \\
(3, 1), (2, 1) & (3, 1)(3, 1) & \dots & (3, 1), (4, 3) \\
(3, 2), (2, 1) & (3, 2)(3, 1) & \dots & (3, 2), (4, 3) \\
(4, 1), (2, 1) & (4, 1)(3, 1) & \dots & (4, 1), (4, 3) \\
(4, 2), (2, 1) & (4, 2)(3, 1) & \dots & (4, 2), (4, 3) \\
(4, 3), (2, 1) & (4, 3)(3, 1) & \dots & (4, 3), (4, 3).
\end{array} \tag{89}$$

Die Bedeutung dieses ansonsten wenig bekannten Produktes ergibt sich aus folgendem

**Satz 64**

Die Matrix  $A$  besitze die Eigenwerte  $\mu_1, \mu_2, \dots, \mu_n$ . Dann hat die Matrix  $A \odot A$  die Eigenwerte  $\mu_i \mu_k$  und die Matrix  $2A \odot I$  die Eigenwerte  $\mu_i + \mu_k$ , wobei  $i = 2(1)n$ ,  $k = 1(1)i - 1$ , und  $I$  die Einheitsmatrix ist.

Setzen wir zur Bestimmung von  $B := 2A \odot I$  die Einheitsmatrix  $I = (\delta_{ik})$  in die Definition ein, so liefert (88) elementeweise

$$B_{(p,q),(r,s)} = \begin{vmatrix} a_{pr} & a_{ps} \\ \delta_{qr} & \delta_{qs} \end{vmatrix} + \begin{vmatrix} \delta_{pr} & \delta_{ps} \\ a_{qr} & a_{qs} \end{vmatrix} = a_{pr}\delta_{qs} - a_{ps}\delta_{qr} + a_{qs}\delta_{pr} - a_{qr}\delta_{ps},$$

mit dem Kronecker-Symbol  $\delta_{ik}$ , woraus sich die einfache Fallunterscheidung

$$B_{(p,q),(r,s)} = \begin{cases} -a_{ps} & \text{für } r = q \\ a_{pr} & \text{für } r \neq p \text{ und } s = q \\ a_{pp} + a_{qq} & \text{für } r = p \text{ und } s = q \\ a_{qs} & \text{für } r = p \text{ und } s \neq q \\ -a_{qr} & \text{für } s = p \\ 0 & \text{sonst} \end{cases} \tag{90}$$

ergibt. Wir können damit die Testfunktion (87) für Hopf-Bifurkation in der Darstellung

$$\psi_H(y) = \det(2D_x f(y) \odot I) \tag{91}$$

angeben, die das Produkt aller Eigenwerte von  $2D_x f(y) \odot I$  liefert.

**Beispiel 65**

$$A = D_x f(y^*) = \begin{pmatrix} 2 & 2 & 7 \\ -5 & -5 & -7 \\ -2 & 5 & 0 \end{pmatrix} \quad \text{mit Eigenwerten } \mu_{1,2} = \pm 7i, \mu_3 = -3$$

sei exemplarisch eine Jacobi-Matrix am Bifurkationspunkt  $y^* = (x^*, \lambda^*)$ . Wegen  $\operatorname{Re} \mu_{1,2} = 0$  und  $\operatorname{Re} \mu_3 < 0$  kann hier eine Hopf-Bifurkation vorliegen. Andererseits kann keine der anderen genannten Bifurkationen eintreten, da  $\psi_F(y^*) = \det A = -147 \neq 0$  ist. Wir berechnen mittels (90) die von  $A$  abgeleitete Matrix

$$B = 2A \odot I = \begin{pmatrix} -3 & -7 & -7 \\ 5 & 2 & 2 \\ 2 & -5 & -5 \end{pmatrix} \quad \text{mit Eigenwerten } \beta_{1,2} = -3 \pm 7i, \beta_3 = 0$$

und erhalten daraus den Wert der Testfunktion  $\psi_H(y^*) = \det B = 0$ . ■

Für große Dimension  $n$  der DGL ist der numerische Aufwand  $T(n)$  zur Determinantenberechnung von  $B = 2A \odot I$  allerdings beträchtlich. Nutzt man die LU-Zerlegung der  $m \times m$ -Matrix  $B$ , so ergibt sich wegen  $m = n(n-1)/2$  asymptotisch

$$T(n) \approx \frac{1}{3}m^3 \approx \frac{1}{3} \left( \frac{1}{2}n^2 \right)^3 \approx \frac{1}{24}n^6,$$

weshalb dann schnelle Approximationen des Spektrums der Jacobi-Matrix  $A$  geeigneter erscheinen. Algorithmen zur Berechnung derartiger *Ritz-Werte* und *Petrov-Werte* großdimensionaler Matrizen findet man u. a. bei Hoffmann et al. (2005) in Kapitel 20.

## 5 Zusammenfassung

Um Gleichgewichtslösungen nichtlinearer dynamischer Systeme – beschrieben durch gewöhnliche Differenzialgleichungen – berechnen und analysieren zu können, wurden folgende Verfahren vorgestellt:

- Das Newton-Verfahren und seine Modifikationen (Sehnenverfahren, Shamanskii-Verfahren, Newton-ähnliche Verfahren, gedämpfte Newton-Verfahren) bilden die Basis für eine effiziente Berechnung von Gleichgewichten als Lösungen nichtlinearer Gleichungssysteme.
- Hängen die Gleichgewichte von einem skalaren Parameter ab, so werden Prädiktor-Korrekturverfahren zur Lösungsfortsetzung eingesetzt. Während die natürliche Parametrisierung mit geeigneter Steuerung der Fortsetzungsschrittweite im regulären Fall gute Ergebnisse liefert, sind beim Auftreten von Umkehrpunkten eine Pseudo-Bogenlängen-Parametrisierung oder die Gauß-Newton-Fortsetzung vonnöten.
- Eine wesentliche Beschleunigung der Parameterfortsetzung kann in vielen Fällen durch die eingeführten Ableitungsprädiktoren erreicht werden. Die Idee des Prädiktor-Korrektor-Verfahrens lässt sich auch zur Fortsetzung mehrdimensionaler Mannigfaltigkeiten nutzen, allerdings kann der Aufwand beträchtlich werden.

- Berechnet man parallel zur Fortsetzung die charakteristischen Exponenten und damit die Dimensionen der stabilen, instabilen und Zentrumsunterräume, so wird die Detektierung lokaler Bifurkationen möglich. Durch Einführung skalarer Testfunktionen für die auftretenden Bifurkationstypen kann insbesondere bei Vorliegen großdimensionaler Systeme die ständige Berechnung der Matrixeigenwerte vermieden werden.

Die vorgestellten Beispiele und nachfolgenden Aufgaben sollen der besseren Veranschaulichung der theoretischen Aussagen dienen.

## 6 Aufgaben

### Aufgabe 1

Betrachten Sie für den Parameter  $s \in \mathbb{R}$  folgendes Differenzialgleichungssystem:

$$\begin{aligned}\dot{x}_1 &= s \cdot x_1 - x_2 - x_1 \cdot (x_1^2 + x_2^2) \\ \dot{x}_2 &= x_1 + s \cdot x_2 - x_2 \cdot (x_1^2 + x_2^2).\end{aligned}$$

- Zeichnen Sie ein Phasenportrait dieses Systems für  $[-2, 2] \times [-2, 2]$ .
- Bestimmen Sie sämtliche Gleichgewichtslösungen (Ruhelagen, stationäre Lösungen) des Differenzialgleichungssystems. Charakterisieren Sie diese Punkte nach Hyperbolizität und Senke-Sattel-Quelle-Eigenschaft.

### Aufgabe 2

Untersuchen Sie sämtliche Gleichgewichtslösungen der Differenzialgleichung

$$\dot{x} = x \cdot (x^3 - x - \lambda)$$

mit Parameter  $\lambda \in \mathbb{R}$  daraufhin, ob sie hyperbolisch, Sättel, Quellen oder Senken sind! Welcher Zusammenhang besteht zum Regularitätsbegriff?

### Aufgabe 3

Bestimmen Sie sämtliche Lösungen der Gleichung  $f(x, \lambda) = 0$  zu gegebenem Parameter  $\lambda \in \mathbb{R}$ , und entscheiden Sie, ob diese regulär oder singulär sind. Dabei ist jeweils  $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ .

- $f(x, \lambda) = \begin{pmatrix} x_1^2 - \lambda^2 \\ x_2^2 + \lambda^2 - 1 \end{pmatrix}$  für  $\lambda = 0, \frac{1}{2}, 1$ .
- $f(x, \lambda) = \begin{pmatrix} 2x_1^2 - 4 + \lambda \cdot (x_1 + x_2) \\ 2x_2^2 - 4 + \lambda \cdot (x_1 + x_2) \end{pmatrix}$  für beliebiges  $\lambda \in \mathbb{R}$ .

### Aufgabe 4

Beweisen Sie, dass unter den Voraussetzungen des Satzes 24

- der klassische Prädiktor die Approximationsordnung 1,

- b) der Tangenten-Prädiktor die Approximationsordnung 2,  
 c) der Sekanten-Prädiktor die Approximationsordnung 1 besitzt.

### Aufgabe 5

Entwickeln Sie in Analogie zu Formel (27) einen Taylor-Prädiktor, der den Prädiktorpunkt  $x^P$  bei Kenntnis der 1. und 2. Ableitung  $x'(\lambda_{j-1})$  und  $x''(\lambda_{j-1})$  bei  $\lambda = \lambda_{j-1}$  ermittelt. Geben Sie Berechnungsvorschriften für  $x'(\lambda_{j-1})$  und  $x''(\lambda_{j-1})$  an, und bestimmen Sie die Approximationsordnung dieser Prädiktorformel. Bewerten Sie den numerischen Aufwand dieser Formel!

### Aufgabe 6

Entwickeln Sie eine MATLAB-Funktion `continuation` für Algorithmus 28 zum Prädiktor-Korrektor-Verfahren mit klassischem Prädiktor bei natürlicher Parametrisierung.

### Aufgabe 7

Zur Bestimmung der Quadratwurzel  $x = \sqrt{\lambda}$ ,  $\lambda > 0$  soll das Prädiktor-Korrektor-Verfahren auf das Nullstellenproblem

$$f(x, \lambda) = x^2 - \lambda = 0, \quad a \leq \lambda \leq b,$$

angewendet werden.

- a) Notieren Sie die Formeln für den Tangenten-Prädiktor und den Newton-Korrektor.  
 b) Starten Sie mit  $\lambda_0 = a = 1$ ,  $x_0 = 1$ , und bestimmen Sie numerisch für  $1 \leq \lambda \leq 5$  die Werte  $x = \sqrt{\lambda}$  mit dem Fortsetzungsverfahren und den Schrittweiten  $s = 1$ ,  $s = 2$ ,  $s = 4$ .  
 c) Vergleichen Sie die Effizienz der drei Rechnungen, und versuchen Sie, die getroffene Aussage zu verallgemeinern.

### Aufgabe 8

- a) Ermitteln Sie alle nichtnegativen Gleichgewichtslagen  $x = (x_1, x_2)^T$ , d. h. mit  $x_1, x_2 \geq 0$ , des DGL-Systems

$$\dot{x}_1 = x_1 \cdot (s - x_1^2), \quad \dot{x}_2 = x_2 \cdot (\lambda - (x_2^2 - 1)^2)$$

mit festem reellen  $s > 0$  und Parameter  $\lambda \in [-2, 2]$ .

- b) Geben Sie die regulären Lösungspfade  $\mathcal{L}$  an!  
 c) Bestimmen Sie alle nicht-regulären Lösungen  $x$ , und prüfen Sie, ob ein Umkehrpunkt vorliegt!

### Aufgabe 9

Welche sind die regulären Lösungspfade der Gleichgewichtslösungen von

$$\dot{x} = x \cdot (x^3 - x - \lambda), \quad \lambda \in [-1, 1] ?$$

Verifizieren Sie, welche Kurvenpunkte regulär und welche Umkehrpunkte sind.

**Aufgabe 10**

Untersuchen Sie die Gleichgewichtslage  $y^0 = (x_1^0, x_2^0, \lambda_0) = (0, 0, 0)$  des Systems

$$\dot{x}_1 = x_1^p - \lambda, \quad \dot{x}_2 = x_2^q - \lambda \quad \text{mit } p, q \in \mathbb{N}$$

daraufhin, ob sie regulär, singular bzw. Umkehrpunkt ist. Führen Sie dabei notwendige Fallunterscheidungen durch!

**Aufgabe 11**

Für die Pseudo-Bogenlängen-Parametrisierung der Kurve  $y = y(s)$  nach Algorithmus 34 wird der normierte Tangentenvektor  $y'(s_j)$  im Punkt  $s_j$  benötigt.

- Geben Sie eine Methode zur Berechnung von  $y'(s_j)$  mit dem richtigen Vorzeichen an, die die LU-Zerlegung benutzt!
- Entwickeln Sie eine MATLAB-Funktion `pseudoarclength` zur Fortsetzung mit Pseudo-Bogenlängen-Parametrisierung gemäß Algorithmus 34 unter Benutzung des Tangenten-Prädiktors und des vereinfachten Newton-Korrektors.
- Erproben Sie diese Funktion anhand der Aufgabe 8.

**Aufgabe 12**

Das DGL-System von L. Chua wird im  $\mathbb{R}^3$  durch

$$\begin{aligned} \dot{x} &= \alpha \cdot (y - x - f(x)) \\ \dot{y} &= x - y + z \\ \dot{z} &= -\beta \cdot y \end{aligned} \quad \text{mit } f(x) = \begin{cases} b \cdot x + a - b & \text{für } x \geq 1 \\ a \cdot x & \text{für } |x| < 1 \\ b \cdot x - a + b & \text{für } x \leq -1 \end{cases}$$

und Parametern  $\alpha, \beta, a, b \in \mathbb{R} \setminus \{0\}, b \neq -1$ , beschrieben.

- Bestimmen Sie alle drei Gleichgewichtslösungen im allgemeinen Fall!
- Im Folgenden seien die Werte  $(\alpha, \beta, a, b) = (9, 14\frac{2}{7}, -\frac{8}{7}, -\frac{5}{7})$  gesetzt. Berechnen Sie damit die drei Gleichgewichtslösungen. Sind diese stabil? Bestimmen Sie die Eigenwerte  $\mu_1, \mu_2, \mu_3$  der Jacobi-Matrizen. Wie groß sind die Dimensionen der Unterräume  $E^s, E^u$  und  $E^c$ ?
- Ermitteln Sie die invarianten Unterräume  $E^s, E^u, E^c$ ! Skizzieren Sie im  $\mathbb{R}^3$  die Ruhpunkte, die invarianten Unterräume und die trennenden Ebenen

$$U_1 = \{(x, y, z) \in \mathbb{R}^3 | x = 1\} \quad \text{und} \quad U_{-1} = \{(x, y, z) \in \mathbb{R}^3 | x = -1\}.$$

Wie verhalten sich die Orbits des Systems?

- Stellen Sie die Orbits – etwa mittels MAPLE – graphisch dar!

**Aufgabe 13**

Entwickeln Sie eine MATLAB-Funktion `gaussnewton` zur Gauß-Newton-Fortsetzung gemäß Algorithmus 39 unter Benutzung des Sekanten-Prädiktors und des (vereinfachten) Gauß-Newton-Korrektors.

**Aufgabe 14**

Das Differenzialgleichungssystem

$$\dot{x}_1 = \varepsilon \cdot x_1 + x_1^2$$

$$\dot{x}_2 = x_2$$

$$\dot{x}_3 = -x_3$$

besitzt für alle  $\varepsilon \in \mathbb{R}$  die instabile Gleichgewichtslage  $x^0 = 0$  (Begründung). Kann  $\varepsilon_0 = 0$  dennoch ein Bifurkationspunkt sein? Entscheiden Sie gegebenenfalls, welcher Bifurkationstyp vorliegt.

**Aufgabe 15**

Die Dynamik eines einfachen Jäger-Beute-Ökosystems mit skalierten Populationszahlen  $x_1 > 0$  und  $x_2 > 0$  wird durch die Differenzialgleichungen

$$\dot{x}_1 = x_1 \cdot (\lambda + x_1) \cdot (1 - x_1) - 3x_1 \cdot x_2$$

$$\dot{x}_2 = -2\lambda \cdot x_2 + x_1 \cdot x_2$$

beschrieben, wobei  $\lambda > 0$  ist.

- Welche Gleichgewichtslösungen besitzt das System? Sind sie stabil?
- Für welche Parameterwerte  $\lambda_0$  wird das betrachtete Gleichgewicht instabil? Welche Bifurkationen treten auf?
- Für welche Parameter besitzt das System eine periodische Lösung? Ist sie stabil? (Begründung!)
- Skizzieren Sie die Phasenportraits für die verschiedenen Lösungen!

**Aufgabe 16**

Programmieren Sie in MATLAB eine Funktion  $B = \text{bialt}(A)$  zur Bestimmung des bialternierenden Produkts  $B = 2A \odot I$  der Matrix  $A \in \mathbb{R}^{n \times n}$ .

- Überprüfen Sie anhand der Testmatrizen  $A = \text{magic}(n)$ ,  $n = 3, 4, 5, \dots$  aus MATLAB und selbst gewählter Matrizen  $A$  die Richtigkeit des Programms.
- Verifizieren Sie mit den Testmatrizen aus MATLAB

$$A = \text{magic}(n) - \text{magic}(n)^T, \quad n = 3, 4, \dots$$

und den Funktionen  $\text{eig}(A)$ ,  $\text{eig}(B)$ ,  $\text{det}(B)$  die Aussagen des Satzes 64 über die Eigenwerte von  $B$ .

- Vergleichen Sie mit selbst gewählten  $n$  und Testmatrizen  $A$  den Rechenaufwand, gemessen in der Zahl der Gleitpunkt-Operationen (Flops), zur Bestimmung aller Eigenwerte über  $\text{eig}(A)$  mit der Zahl der Flops bei  $\text{det}(\text{bialt}(A))$ . Versuchen Sie eine theoretische Aussage zur Zeitkomplexität  $T(n)$  für beide Varianten.

# Literaturverzeichnis

- Allgower E. L. und Georg K. (1990) *Numerical Continuation Methods*. Springer-Verlag, Berlin.
- Argyris J., Faust G. und Haase M. (1995) *Die Erforschung des Chaos*. Vieweg Verlag, Braunschweig und Wiesbaden.
- Deuffhard P. und Hohmann A. (1993) *Numerische Mathematik I*. 2. Auflage, W. de Gruyter, Berlin.
- Deuffhard P. (2004) *Newton Methods for Nonlinear Problems*. Springer-Verlag, Berlin.
- Doedel E. J., Paffenroth R. C., Champneys A. R., Fairgrieve T. F., Kuznetsov Y. A., Oldeman B. E., Sandstede B. und Wang X. (2002) *AUTO 2000 - Continuation and bifurcation software for ordinary differential equations (with HomCont), users guide*. Concordia University, Montreal, P.Q., Canada.
- Henderson M. E. (2007) *Higher-Dimensional Continuation*. In: Krauskopf B., Osinga H.M. und Galán-Vioque J. (Hrsg.) *Numerical Continuation Methods for Dynamical Systems*. Springer-Verlag Netherlands.
- Hoffmann A., Marx B. und Vogt W. (2005) *Mathematik für Ingenieure. Band 1: Lineare Algebra, Analysis – Theorie und Numerik*. Pearson Studium, München.
- Hoffmann A., Marx B. und Vogt W. (2006) *Mathematik für Ingenieure. Band 2: Vektoranalysis, Integraltransformationen, Differenzialgleichungen, Stochastik – Theorie und Numerik*. Pearson Studium, München.
- Isaacson E. und Keller H. B. (1972) *Analyse numerischer Verfahren*. Verlag Harry Deutsch, Frankfurt.
- Kelley C. T. (1995) *Iterative Methods for Linear and Nonlinear Equations*. SIAM Publications, Philadelphia.
- Kosmol P. (1993) *Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben*. B.G.Teubner, Stuttgart.
- Kuznetsov Y. A. (1995) *Elements of Applied Bifurcation Theory*. Springer-Verlag, New York.
- Langford W. F. (1984) *Numerical Studies of Torus Bifurcations*. In: International Series of Numerical Mathematics, Vol.70, Birkhäuser Verlag, Basel.
- Marx B. und Vogt W. (2010) *Dynamische Systeme – Theorie und Numerik*. Erscheint bei Spektrum-Verlag, Heidelberg.
- Mei Zhen (2000) *Numerical Bifurcation Analysis for Reaction-Diffusion Equations*. Springer-Verlag, Berlin.
- Rheinboldt W. C. (1994) *Methods for Solving Systems of Nonlinear Equations*. 4th ed., SIAM Publications, Philadelphia.
- Schwetlick H. (1979) *Numerische Lösung nichtlinearer Gleichungen*. Deutscher Verlag der Wissenschaften, Berlin.